

MASTER'S THESIS

Het uitvoeren van een experiment om de intuïtiviteit van verschillende versies van de PGA-notatie vergelijkend te evalueren

van den Oever, M. (Merijn)

Award date:
2021

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 05. May. 2023

Open Universiteit
www.ou.nl



Het uitvoeren van een experiment om de intuïtiviteit van verschillende versies van de PGA-notatie vergelijkend te evalueren

Performing an experiment to comparatively evaluate the intuitiveness of different versions of the PGA notation

Opleiding:	Open Universiteit, faculteit Management, Science & Technology Masteropleiding Business Process Management & IT
Programma:	Open University of the Netherlands, faculty of Management, Science & Technology Master Business Process Management & IT
Cursus:	IM0602 Voorbereiden Afstuderen BPMIT IM9806 Afstudeeropdracht Business Process Management and IT
Student:	Merijn van den Oever
Identiteitsnummer:	
Datum:	24-01-2021
Afstudeerbegeleider	Ben Roelens
Meelezer	Harry Martin
Derde beoordelaar	n.v.t.
Versie nummer:	1.0
Status:	Definitief

Abstract

De Process-Goal Alignment (PGA) notatie is een domein specifieke modelleertaal (DSML) die tot doel heeft om strategic fit te bereiken in de bedrijfsarchitectuur en hiertoe inzichtelijk maakt in welke mate bedrijfsactiviteiten de bedrijfsstrategie ondersteunen. Om adoptie en een correct gebruik ervan te verzekeren is het van belang dat de modelleertaal intuïtief is en daardoor eenvoudig door de gebruiker is te interpreteren. Om de intuïtiviteit te verbeteren zijn middels een evaluatietechniek verbeteringen op de notatie doorgevoerd. Echter, tot nu toe is niet met zekerheid te zeggen of deze techniek tot significante verbeteringen heeft geleid voor de daadwerkelijke eindgebruikers. Om deze research gap te overbruggen is in dit onderzoek een vervolgonderzoek opgezet met notatieassociatie-, begrips-, herinnerings- en representatievoorkeur taken die de intuïtieve begrijpelijkheid vergelijkend evalueert tussen twee versies van de PGA-notatie in een praktische bedrijfscontext. De intuïtiviteit is gemeten met behulp van de objectieve variabelen interpretatieve effectiviteit, interpretatieve efficiëntie en de subjectieve variabele representatievoorkeur. De onderzoeksresultaten van de representatievoorkeur taak laten zien dat de evaluatietechniek heeft geleid tot subtiele verbeteringen voor eindgebruikers, die door de overige taken niet zijn gedetecteerd. Hiertoe zijn concrete voorstellen gedaan voor verbeteringen in de validiteit en betrouwbaarheid van de experimentele opzet die een startpunt vormen voor vervolgonderzoek.

Sleutelbegrippen

Process-Goal Alignment (PGA) notatie, Domain Specific Modeling Language (DSML),
intuïtieve begrijpelijkheid, representatievoorkeur, experimentele evaluatie, empirisch vergelijken

Samenvatting

Conceptuele domein specifieke modelleertalen (DSML's) zijn bedoeld voor bepaalde toepassingsdomeinen en bevatten modelleerconcepten waarmee eindgebruikers bekend zijn. Specifiek voor DSML's speelt de intuïtiviteit (i.e. het gemak waarmee gebruikers het diagram direct kunnen begrijpen, zonder enige voorkennis of training) van de notatie een belangrijke rol. Hierdoor wordt zowel het modelleren als het interpreteren van het model vergemakkelijkt waardoor de kans op adoptie en een correct gebruik ervan wordt vergroot.

De Process-Goal Alignment (PGA) notatie - die tot doel heeft om strategic fit te bereiken in de bedrijfsarchitectuur - is een DSML waarbij intuïtiviteit belangrijk is en waarvoor middels een evaluatietechniek verbeteringen op de notatie zijn doorgevoerd op vijf elementen. Echter, tot nu toe is niet met zekerheid te zeggen of deze techniek tot significante verbeteringen heeft geleid voor de daadwerkelijke eindgebruikers (i.e. business users). Daarom is in dit onderzoek op basis van methodologische literatuur een vervolgonderzoek opgezet om met behulp van een vergelijkende evaluatie de intuïtieve begrijpelijkheid te evalueren tussen twee versies van de PGA-notatie (i.e. 'nieuwe' versus 'originele' versie) in een praktische bedrijfscontext. In het experiment is met een viertal taken (i.e. notatieassociatie-, begrips-, herinnerings- en representatievoorkeur taak) de intuïtiviteit gemeten voor vijf elementen (i.e. value proposition, customer goal, competence, internal goal en valuestream/importance). Hierbij is gebruik gemaakt van de objectieve variabelen interpretatieve effectiviteit (i.e. het aantal correcte antwoorden op de notatieassociatie-, begrips-, herinneringstaken), interpretatieve efficiëntie (i.e. de tijd in seconden die de participant nodig heeft om de notatieassociatie-, begrips-, herinneringstaken af te ronden), de subjectieve variabele representatievoorkeur (i.e. naar welk versie van de PGA-notatie de intuïtieve voorkeur uitgaat bij het zien van beide notaties in de representatievoorkeur taken), en de geaggregeerde variabele voor het hele construct intuïtieve begrijpelijkheid (i.e. geaggregeerde score van interpretatieve effectiviteit, - efficiëntie en representatievoorkeur). Vervolgens is met behulp van een online survey een experiment uitgevoerd onder 15 participanten (i.e. business users in een praktische bedrijfscontext) die in een tijdsperiode van drie weken kon worden ingevuld. Hierbij zijn de participanten automatisch in twee subgroepen verdeeld met elk een eigen survey waarin de notaties zijn gewisseld en voor alle taken dezelfde combinatie van 'originele' en 'nieuwe' notaties wordt gebruikt. Om de kracht van de evaluatieresultaten te vergroten zijn de data van twee separate onderzoeken die tegelijkertijd zijn uitgevoerd bij andere casusorganisaties (i.e. 29 participanten) ook in het onderzoek meegenomen, waardoor analyse heeft plaatsgevonden op zowel de eigen (i.e. 15 participanten) als de totale (i.e. 44 participanten) dataset.

Van de vijf elementen van de PGA-notatie die door de evaluatietechniek zijn verbeterd, is volgens dit onderzoek voor twee elementen (i.e. value proposition en customer goal) de 'nieuwe' notatie als significant beter beoordeeld in vergelijking met de 'originele'. Dit is vastgesteld met de representatievoorkeur taak, het enige meetinstrument waarvan de constructvaliditeit als voldoende kan worden bestempeld. Hoewel dit in vervolgonderzoek moet worden gevalideerd, kan dus voorlopig worden aangenomen dat de intuïtiviteit van verschillende versies van een DSML-notatie empirisch met elkaar kan worden vergeleken, door gebruik te maken van de representatievoorkeur variabele, die kan worden gemeten met de representatievoorkeur taak waarin zowel kwantitatieve als kwalitatieve data worden verzameld. Op basis hiervan kan vervolgens middels triangulatie worden vastgesteld in welke mate de kwalitatieve data de resultaten van de statistische tests op de kwantitatieve data ondersteunen. Niet geschikt voor het detecteren van verbeteringen zijn de in het onderzoek gebruikte meetinstrumenten voor interpretatieve effectiviteit en -efficiëntie: de

notatieassociatie-, begrips- en herinneringstaak. Dit lijkt een opvallend gegeven, aangezien de evaluatietechniek waarmee de verbeteringen zijn aangebracht deels gebruik maakt van dezelfde taken (i.e. de notatieassociatie- en begripstaak) en slechts lijkt te hebben geleid tot subtiele verbeteringen die door eindgebruikers in een andere taak (i.e. representatievoorkeur taak) deels als significant beter worden beoordeeld. Daarom is het in het belang van de evaluatietechniek én de vervolgevaluatie van dit onderzoek noodzakelijk om in vervolgonderzoek naar meetinstrumenten te zoeken die beter in staat zijn om de intuïtiviteit (i.e. interpretatieve effectiviteit en -efficiëntie) te meten, of de reeds gebruikte meetinstrumenten fijner af te stellen, zodat deze in staat zijn om meer subtiele verbeteringen te meten. Ook is niet zeker of de variabelen interpretatieve effectiviteit en -efficiëntie wel geschikt zijn voor het meten van intuïtieve begrijpelijkheid, aangezien hierover in dit onderzoek geen uitspraak kon worden gedaan. Om die reden moet hier in vervolgonderzoek extra aandacht aan worden besteed en wellicht naar andere variabelen worden gezocht die hier wel toe in staat zijn. Aangezien de drie separate onderzoeken niet tot dezelfde onderzoeksresultaten hebben geleid, kan worden geconcludeerd dat de betrouwbaarheid als matig kan worden bestempeld. Een herhaling van het onderzoek zal dan waarschijnlijk ook niet leiden tot dezelfde resultaten. Daarom is het aan te raden om in vervolgonderzoek vooraf goed stil te staan bij de modelleerervaring en -expertise; eigenschappen die participanten zouden moeten hebben in relatie tot de modelleertaal die wordt geëvalueerd. Er zijn immers aanwijzingen dat onderlinge verschillen hierin tussen participanten een verstoring effect hebben gehad op de onderzoeksresultaten. Participanten van de eigen dataset beoordelen de nieuwe notatie namelijk minder vaak (i.e. 3 keer) als significant beter in vergelijking met de totale dataset (i.e. 6 keer); iets wat mogelijk kan worden verklaard door verschillen in modelleerexpertise- en ervaring. Participanten van het eigen onderzoek hadden namelijk in vergelijking met de andere datasets aanzienlijk vaker geen enkele modelleertaal ervaring; minder ervaring met BPM-modelleertalen; en vaker geen enkele modelleerexpertise. Deze eigenschappen zouden vooraf vastgesteld moeten worden door onderzoek hiernaar te doen bij daadwerkelijke eindgebruikers en vervolgens worden gebruikt als inclusie-/exclusie criteria bij het selecteren van participanten.

In de praktijk kunnen organisaties in de toekomst mogelijk gebruik maken van een meer fijn geslepen versie van de evaluatietechniek voor het verbeteren van DSML's, waarbij gebruik moet worden gemaakt van participanten met de specifieke eigenschappen (i.e. modelleerexpertise en -ervaring) van de daadwerkelijke eindgebruikers. De verbeteringen kunnen vervolgens met een verbeterde versie van de vervolgevaluatie worden beoordeeld door een andere groep participanten. Om er nadien zeker van te zijn dat de verbeteringen eveneens als beter worden beoordeeld door de bestaande gebruikers is het van belang om daadwerkelijke eindgebruikers de potentiële verbeteringen te laten valideren. Wanneer voldoende vertrouwen is opgebouwd met de vervolgevaluatie, dan is deze extra stap wellicht niet meer nodig.

Op basis van bovenstaande kan worden geconcludeerd dat het onderzoek een bijdrage levert aan de body of knowledge over het empirisch evalueren van de intuïtiviteit van DSML-notaties. Hoewel voor verbetering vatbaar levert het namelijk een startpunt voor een methode die de intuïtieve begrijpelijkheid van verschillende versies van een DSML-notatie empirisch met elkaar vergelijkt. Ook is het onderzoek - in tegenstelling tot voorgaand onderzoek - gedaan in een praktische bedrijfscontext, waardoor inzichtelijk is geworden dat mogelijk extra aandacht nodig is bij het selecteren van participanten (i.e. modelleerexpertise en -ervaring) voor het evalueren van DSML's. De laatste bijdrage heeft van doen met het feit dat een eerste inzicht is verkregen in hoezeer de evaluatietechniek voor het verbeteren van DSML-notaties tot significante verbeteringen heeft geleid voor de daadwerkelijke eindgebruikers.

Summary

Conceptual domain specific modeling languages (DSMLs) are intended for particular application domains and contain modeling concepts that end users are familiar with. Specifically for DSMLs, the intuitiveness (i.e. the ease with which users can understand the diagram directly, without any prior knowledge or training) of the notation plays an important role. This facilitates both modeling and interpretation of the model, increasing the chance of adoption and correct use.

The Process-Goal Alignment (PGA) notation - which aims to achieve strategic fit in the business architecture - is a DSML in which intuitiveness is important and for which improvements have been made to the notation on five elements by means of an evaluation technique. However, it is not yet possible to say with certainty whether this technique has led to significant improvements for the actual end users (i.e. business users). Therefore, in this study, based on methodological literature, a follow-up experiment was set up to evaluate, using a comparative evaluation, the intuitive comprehensibility between two versions of the PGA notation (i.e. "new" versus "original" version) in a practical business context. In the experiment, the intuitiveness was measured for five elements (i.e. value proposition, customer goal, competence, internal goal and value stream / importance) using four tasks (i.e. notation association, comprehension, recall and representation preference task). Use was made of the objective variables interpretative effectiveness (i.e. the number of correct answers to the notation association, comprehension and recall tasks), interpretative efficiency (i.e. the time in seconds that the participant needs to complete the notation association, comprehension and recall tasks), the subjective variable representation preference (i.e. which version of the PGA notation is the intuitive preference when seeing both notations in the representation preference tasks), and the aggregated variable for the whole construct intuitive comprehensibility (i.e. aggregated score of interpretive effectiveness, efficiency and representation preference). Subsequently, an experiment was conducted among 15 participants (i.e. business users in a practical business context) using an online survey, which could be completed over a period of three weeks. The participants are automatically divided into two subgroups, each with its own survey in which the notations have been changed and the same combination of "original" and "new" notations is used for all tasks. In order to increase the strength of the evaluation results, the data from two separate studies that were carried out simultaneously at other case organizations (i.e. 29 participants) were also included in the study, as a result of which analysis was carried out on both the own (i.e. 15 participants) and the total (i.e. 44 participants) dataset.

Of the five elements of the PGA notation that have been improved by the evaluation technique, according to this study, for two elements (i.e. value proposition and customer goal) the "new" notation was rated as significantly better compared to the "original". This has been determined with the representation preference task, the only measuring instrument for which the construct validity can be considered satisfactory. While this needs to be validated in further research, it can therefore be provisionally assumed that the intuitiveness of different versions of a DSML notation can be empirically compared with each other, using the representation preference variable, which can be measured with the representation preference task in which both quantitative and qualitative data are collected. Based on this, it can then be determined by means of triangulation to what extent the qualitative data support the results of the statistical tests on the quantitative data. The following measuring instruments for interpretive effectiveness and efficiency used in the study are not suitable for detecting improvements: the notation association, comprehension and recall tasks. This seems a striking fact, as the evaluation technique with which the improvements were made partly uses the same tasks (i.e. the notation association and comprehension task) and only appears to have led to

subtle improvements that end users in another task (i.e. representation preference task) have partly assessed as significantly better. Therefore, in the interest of the evaluation technique and the follow-up evaluation of this research, it is necessary in further research to search for measuring instruments that are better able to measure intuitiveness (i.e. interpretative effectiveness and efficiency), or to fine-tune the already used measuring instruments so that they are able to measure more subtle improvements. Is it also not certain whether the variables interpretative effectiveness and efficiency are suitable for measuring intuitive comprehensibility, as no statement could be made about this in this study. For this reason, additional attention should be paid to this in further research and perhaps other variables that are capable of this should be looked for. Since the three separate studies did not lead to the same study results, it can be concluded that the reliability can be described as moderate. A repetition of the study will therefore probably not lead to the same results. It is therefore advisable to consider modeling experience and expertise in advance in follow-up research; properties that participants should have in relation to the modeling language being evaluated. After all, there are indications that differences in this between participants had a disruptive effect on the research results. Participants of the own dataset assess the new notation less often (i.e. 3 times) as significantly better compared to the total dataset (i.e. 6 times); something that could possibly be explained by differences in modeling expertise and experience. In comparison with the other datasets, participants in the own research significantly more often had no experience in any modeling language; less experience with BPM modeling languages; and more often no modeling expertise. These traits should be predetermined by researching them on actual end users and then be used as inclusion / exclusion criteria when selecting participants.

In practice, in the future, organizations may be able to use a more refined version of the evaluation technique to improve DSMLs, using participants with the specific characteristics (i.e. modeling expertise and experience) of the actual end users. The improvements can then be assessed by another group of participants with an improved version of the follow-up evaluation. In order to ensure that the improvements are also rated as better by existing users, it is important to have actual end users validate the potential improvements. When sufficient confidence has been built up with the follow-up evaluation, this extra step may no longer be necessary.

Based on the above, it can be concluded that the research contributes to the body of knowledge about the empirical evaluation of the intuitiveness of DSML notations. Although there is room for improvement, it provides a starting point for a method that empirically compares the intuitive comprehensibility of different versions of a DSML notation. In contrast to previous research, the research was also conducted in a practical business context, which made it clear that additional attention may be required when selecting participants (i.e. modeling expertise and experience) for evaluating DSMLs. The latter contribution has to do with the fact that a first insight has been obtained into how much the evaluation technique for improving DSML notations has led to significant improvements for the actual end users.

Inhoudsopgave

Abstract	ii
Sleutelbegrippen	ii
Samenvatting	iii
Inhoudsopgave	vii
1. Introductie	1
1.1. Achtergrond	1
1.2. Gebiedsverkenning	2
1.3. Probleemstelling	3
1.4. Opdrachtformulering	3
1.5. Motivatie / relevantie	4
1.5.1. Wetenschappelijke relevantie	4
1.5.2. Maatschappelijke relevantie.....	4
1.6. Aanpak in hoofdlijnen	4
2. Theoretisch kader	5
2.1. Onderzoeksaanpak	5
2.1.1. Zoekstrategieën	5
2.1.2. Building blocks methode.....	5
2.1.3. Forward snowballing methode	8
2.2. Uitvoering.....	8
2.2.1. Resultaten Building Blocks methode	8
2.2.2. Resultaten Forward snowballing methode.....	9
2.3. Resultaten en conclusies.....	9
2.3.1. Maatstaven	10
2.3.2. Experimenteel Design	13
2.3.3. Experimentele Taken	13
2.3.4. Participanten.....	14
2.3.5. Procedures	15
2.3.6. Raamwerk	15
2.3.7. Conclusie	16
2.4. Doel van het vervolgonderzoek	18
3. Methodologie.....	19
3.1. Conceptueel ontwerp: keuze van onderzoeksmethode(n)	19
3.2. Technisch ontwerp: uitwerking van de methode	20

3.2.1.	Variabelen & maatstaven.....	20
3.2.2.	Experimentele taken	20
3.2.3.	Operationele procedures	21
3.2.4.	Experimentele Groepen: Between- vs. Within-Subjects.....	21
3.2.5.	Hypothesen	22
3.2.6.	Participanten	23
3.3.	Gegevensanalyse.....	23
3.4.	Reflectie t.a.v. validiteit, betrouwbaarheid en ethische aspecten	25
3.4.1.	Constructvaliditeit.....	25
3.4.2.	Interne validiteit.....	25
3.4.3.	Externe validiteit	26
3.4.4.	Betrouwbaarheid	26
3.4.5.	Ethische Aspecten	26
4.	Resultaten	27
4.1.	Participanten en experimentele opzet	27
4.1.1.	Participanten	27
4.1.2.	Experimentele opzet	27
4.2.	Analyse resultaten per element op taakniveau	29
4.2.1.	Resultaten eigen dataset	29
4.2.2.	Resultaten volledige dataset.....	29
4.3.	Analyse geaggregeerde resultaten intuïtiviteit per element.....	31
4.4.	Kwalitatieve analyse	32
4.4.1.	Reacties tijdens en na experiment.....	32
4.4.2.	Kwalitatieve feedback representatievoorkeur	32
5.	Discussie, conclusies en aanbevelingen.....	34
5.1.	Discussie onderzoeksresultaten.....	35
5.1.1.	Samenvatting kwantitatieve en kwalitatieve resultaten per element	35
5.1.2.	Value proposition.....	36
5.1.3.	Customer goal	36
5.1.4.	Competence	36
5.1.5.	Internal goal	37
5.1.6.	Value stream/Importance.....	37
5.2.	Reflectie validiteit en betrouwbaarheid	38
5.2.1.	Constructvaliditeit.....	38
5.2.2.	Interne validiteit.....	39
5.2.3.	Externe validiteit	40

5.2.4. Betrouwbaarheid	40
5.2.5. Ethische Aspecten	41
5.3. Conclusies	42
5.4. Aanbevelingen voor de praktijk	44
5.5. Aanbevelingen voor verder onderzoek.....	45
Referenties	47
Bijlage 1: Begrippenlijst	49
Bijlage 2: Gedetailleerde analyse paper selectie theoretisch kader	51
Bijlage 3: Samenvatting papers theoretisch kader	56
Bijlage 4: ‘Originele’ en ‘nieuwe’ PGA-notatie elementen	60
Bijlage 5: Survey met experimentele taken (voor gebruik in Limesurvey).....	61
Bijlage 6: PGA-model voor begripstaken	62
Bijlage 7: Totaaloverzicht hypothesen.....	64
Bijlage 8: Gedetailleerde analyse onderzoeksresultaten per element op taakniveau van de eigen dataset	67
Bijlage 9: Gedetailleerde analyse onderzoeksresultaten per element op taakniveau van de totale dataset	69
Bijlage 10: Gedetailleerde analyse geaggregeerde onderzoeksresultaten intuïtiviteit van de eigen dataset	72
Bijlage 11: Gedetailleerde analyse geaggregeerde onderzoeksresultaten intuïtiviteit van de totale dataset	75
Bijlage 12: Gedetailleerde analyse kwalitatieve feedback representatievoorkeur	78

1. Introductie

1.1. Achtergrond

Volgens Lankhorst (2009) is Enterprise Architectuur (EA) een instrument voor het managen van de dagelijkse bedrijfsactiviteiten en toekomstige ontwikkelingen, dat gedefinieerd kan worden als *“een samenhangend geheel van principes, methoden en modellen die worden gebruikt bij het ontwerpen en realiseren van de organisatiestructuur, bedrijfsprocessen, informatiesystemen en infrastructuur van een organisatie”* (p. 30).

Om de EA te kunnen beschrijven wordt gebruik gemaakt van EA-modelleertalen: conceptuele modellen die – voor begrips- en communicatiedoeleinden – beogen om de realiteit van een specifiek aandachtsgebied minder complex weer te geven, door alleen de hiervoor relevante aspecten te benadrukken (Bork, Karagiannis, & Pittl, 2018). The Open Group (2019, p. 1) omschrijft deze als *“een visuele taal met een set standaardiconen voor het beschrijven, analyseren en communiceren van EA aandachtsgebieden die in de loop van de tijd veranderen”*. Door het beschrijven van EA worden de complexiteit en onderlinge relaties van architecturale domeinen vastgelegd op een zinvol abstractieniveau. Dit helpt bij het analyseren van de impact van veranderingen en faciliteert de communicatie en besluitvorming hierover (Lankhorst, 2009). Een modelleertaal bevat een syntax, bestaande uit elementen en relaties (i.e. vocabulaire) en een set compositieregels (i.e. grammatica), zoals gedefinieerd in het metamodel (Moody, 2009). Daarnaast bevat een taal definities van de syntax binnen het meta-model (i.e. semantiek) en een grafische visualisatie van deze elementen en relaties (i.e. notatie) (Bork, Karagiannis, & Schruffer, 2019).

Afhankelijk van de toepassing kunnen verschillende soorten EA-modelleertalen worden onderscheiden. Sommige modelleertalen zijn gericht op algemene toepasbaarheid en brede acceptatie: General-Purpose Modeling Languages (GPML's) (Bork et al., 2018). GPML's zijn generieke modelleertalen die onafhankelijk zijn van het gebruiksdomein - en daarom geschikt voor een breed scala aan domeinen (Frank, 2010). Voorbeelden van standaardmodelleertalen zijn Unified Modeling Language (UML), Archimate en Business Process Modeling and Notation (BPMN). Andere modelleertalen richten zich juist op een specifiek doel (e.g. analyse en communicatie over een bepaald probleem) in een gebruiksdomein: Domain-Specific Modeling Languages (DSML's) (Roelens & Bork, 2020). DSML's zijn bedoeld voor bepaalde toepassingsdomeinen en bevatten modelleerconcepten die verrijkt zijn met termen en visualisaties waarmee eindgebruikers bekend zijn (Frank, 2010). Hierdoor wordt zowel het modelleren als het interpreteren (i.e. duidelijkheid en begrijpelijkheid) van het model vergemakkelijkt.

Voor de effectiviteit van een DSML speelt de intuïtiviteit (i.e. intuïtieve begrijpelijkheid) van de notatie een belangrijke rol. Deze notatie zou namelijk afgestemd moeten zijn op de kennis, overtuigingen en verwachtingen van de eindgebruikers. De mate waarin modellen eenvoudig door de gebruiker zijn te interpreteren wordt namelijk gezien als een eerste voorwaarde voor adoptie en een correct gebruik ervan (Bork et al., 2019). Deze interpretatie kan worden vergemakkelijkt door de notatie nauw te doen aansluiten bij hoe de mens effectief informatie verwerkt. Informatieverwerking vindt immers op twee manieren plaats: perceptuele verwerking en cognitieve verwerking. Perceptuele verwerking (i.e. zien) is snel en automatisch, cognitieve verwerking (i.e. begrijpen) is langzaam en energie-intensief. Het is voor de intuïtiviteit van een modelleertaal dus belangrijk om zich te richten op de perceptuele

verwerking, waarbij de betekenis van de elementen door eindgebruikers meteen is af te leiden vanuit de visuele weergave.

Overigens is in bijlage 1 voorzien in een begrippenlijst die helpt om in de veelheid van jargon beter de weg te vinden.

1.2. Gebiedsverkenning

Gezien het belang van een intuïtieve modelleertaal is door Bork et al. (2019) een experimentele methode ontwikkeld die de intuïtiviteit van DSML's evalueert en verbetert. De techniek bestaat uit drie fasen: in fase 1 (i.e. termassociatie) tekent de participant één of meer notaties die hij/zij als meest intuïtief beschouwt voor een reeks elementen en relaties uit het meta-model; in fase 2 (i.e. notatieassociatie) geeft de participant maximaal drie intuïtieve benamingen bij een aantal bestaande notaties; in fase 3 (i.e. case study) bouwt de participant met de DSML een model op basis van een casus, of beantwoordt deze begripsvragen over een bestaand model. Tot slot wordt de participant gevraagd om een survey in te vullen met positieve en negatieve feedback, alsook verbeter suggesties voor de intuïtiviteit van DSML-notaties.

Na het experiment worden de resultaten geanalyseerd op basis waarvan de notatie kan worden verbeterd. De ontvangen intuïtieve notaties van fase 1 worden geclassificeerd op basis van een aantal visuele variabelen (e.g. vorm en kleur). Vervolgens wordt per term de meest voorkomende visuele weergave vergeleken met die van de bestaande notatie. De mate waarin beide met elkaar overeenkomen resulteert in mogelijke verbeteringen voor de bestaande notatie. Voor de analyse van de notatieassociatie (i.e. fase 2) moet worden beoordeeld in hoeverre de gegeven benamingen overeenkomen met de daadwerkelijke termen uit het meta-model. Op basis hiervan worden de elementen en relaties ingedeeld in de klassen 'geïdentificeerd', 'deels geïdentificeerd' en 'niet geïdentificeerd'. Vervolgens wordt met een percentage de totaalscore berekend welk percentage van de respondenten de bestaande notatie intuïtief met de juiste naam associeert. Wanneer in fase 3 door de participant modellen zijn gemaakt, dan wordt beoordeeld hoezeer hierin sprake is van 'toepassingsfouten' (i.e. verkeerde toepassing van een concept of verkeerde definitie van concepteigenschappen), 'procedurefouten' (i.e. verkeerde volgorde van concepten en verkeerde/ontbrekende toepassing van een relatie) en/of een 'onvolledig model' (i.e. ontbrekende - eigenschappen van een - concept). Als in fase 3 begripsvragen zijn gesteld bij een gegeven model, dan wordt beoordeeld tot op welke hoogte de vragen correct beantwoord zijn. Uit de resultaten volgt per element een conclusie: 'correct', 'deels correct' of 'niet correct'. De evaluatiecyclus wordt afgesloten door op basis van de gecombineerde onderzoeksresultaten van fase 1, 2, 3 en de survey vast te stellen of notatieverbeteringen nodig zijn. Wanneer de experimentele resultaten de intuïtiviteit van de notatie bevestigen, dan is er geen wijziging vereist. Als dit niet het geval is, dan wordt een nieuwe notatie voorgesteld op basis van de suggesties van participanten.

In voorgaand hoofdstuk is te lezen dat de intuïtiviteit van notaties vooral belangrijk is bij DSML's. De Process-Goal Alignment (PGA) modelleertaal is een voorbeeld van een DSML waarbij duidelijke communicatie met eindgebruikers een ontwerpvereiste is. Deze modelleertaal heeft tot doel om strategisch fit te bereiken in de bedrijfsarchitectuur en maakt met een heat map inzichtelijk in welke mate bedrijfsactiviteiten (i.e. interne infrastructuur en processen) de bedrijfsstrategie ondersteunen (Roelens, Steenacker, & Poels, 2019). De heat map presenteert kritieke waarde stroom relaties tussen bedrijfsarchitectuur elementen op verschillende hiërarchische niveaus (i.e. strategisch, infrastructureel, operationeel), waarmee vervolgens verbeterkansen kunnen worden geïdentificeerd

voor strategic fit. De PGA-modeleer- en analyseprocedure bestaat uit drie activiteiten: (i) het ontwikkelen van een geprioriteerde bedrijfsarchitectuur hiërarchie; (ii) het uitvoeren van een prestatiemeting; en (iii) het uitvoeren van een verbeteringsanalyse (Roelens & Bork, 2020). Eindgebruikers zijn stakeholders binnen de organisatie (i.e. business users in een praktische bedrijfscontext) die belang bij hebben bij het verbeteren van strategic fit en toegang hebben tot de noodzakelijke interne informatie om het model op te kunnen bouwen.

1.3. Probleemstelling

Hoewel de evaluatietechniek op het moment van schrijven slechts twee keer is toegepast, lijkt de techniek bruikbaar en heeft deze geleid tot mogelijke verbeteringen in de intuïtiviteit van de BCM (i.e. Business Continuity Management) (Bork et al., 2019) en PGA (Roelens & Bork, 2020) notaties. Echter, de analyse van de evaluatieresultaten en het ontwerp van de verbeterde notatie worden in beide onderzoeken op een andere manier uitgevoerd en zijn voor een deel afhankelijk van menselijke creativiteit. Dit maakt de kans groot dat de interpretatie van de resultaten en de doorgevoerde verbeteringen beïnvloed worden door en afhankelijk zijn van de onderzoeker. Deze kunnen daardoor als subjectief bestempeld worden. Wat ook opvalt is dat in beide onderzoeken (Bork et al., 2019; Roelens & Bork, 2020) de daadwerkelijke eindgebruikers van de DSML's (i.e. business users) niet overeenkomen met de participanten in het onderzoek (i.e. masterstudenten). Hierdoor is niet met zekerheid te zeggen of de evaluatietechniek tot significante verbeteringen heeft geleid voor de daadwerkelijke eindgebruikers.

Verder experimenteel onderzoek is dus noodzakelijk om vast te kunnen stellen in hoeverre de evaluatietechniek van Bork et al. (2019) leidt tot significante verbeteringen in de intuïtiviteit van een DSML-notatie voor eindgebruikers in een praktische bedrijfscontext. Dit zal worden gedaan door het opzetten en uitvoeren van een methode, waarbij een vergelijking wordt gemaakt tussen de intuïtiviteit van een 'originele' DSML-notatie en de op basis van de evaluatietechniek verbeterde 'nieuwe' notatie. Dit zal worden gedaan voor de DSML die door Roelens & Bork (2020) reeds met de evaluatietechniek is geëvalueerd: de PGA-notatie.

Samenvattend kan de centrale onderzoeksvraag als volgt worden geformuleerd: *Hoe kan de intuïtiviteit van verschillende versies van een DSML-notatie empirisch met elkaar worden vergeleken in een praktische bedrijfscontext?*

1.4. Opdrachtformulering

Het doel van het onderzoek is het opzetten en uitvoeren van een methode om de intuïtiviteit van verschillende versies van een DSML-notatie empirisch met elkaar te kunnen vergelijken. Om deze vergelijking te kunnen maken zal middels literatuuronderzoek worden geïnventariseerd welke bestaande methoden hiervoor relevant zijn. Hiermee zal een methode worden opgezet die de intuïtiviteit van de 'originele' en 'nieuwe' PGA-notatie empirisch met elkaar vergelijkt. Belangrijk hierbij is het beschrijven van de specifieke bedrijfscontext waarin het onderzoek plaatsvindt met relevante details over de casusorganisatie en de respondenten. Ook wordt beschreven hoe het onderzoek precies zal worden uitgevoerd en wordt stilgestaan bij zaken rondom validiteit en betrouwbaarheid. Daarna zal de methode worden uitgevoerd bij een tiental respondenten en vindt op basis van de onderzoeksresultaten een analyse plaats op de intuïtiviteit van de PGA-notaties. Tot slot zal een conclusie worden getrokken in hoeverre de 'nieuwe' PGA-notatie intuïtiever is dan de

‘originele’ en zal daarnaast ook een uitspraak worden gedaan over de mate waarin de evaluatietechniek van Bork et al. (2019) heeft geleid tot significante verbeteringen in de intuïtiviteit van de PGA-notatie voor eindgebruikers in een praktische bedrijfscontext.

Samenvattend zal een antwoord gegeven worden op de volgende deelvragen die samen tot een antwoord op de centrale onderzoeksvraag gaan leiden:

1. *Welke bestaande methoden zijn relevant om de intuïtiviteit van conceptuele modellen (i.e. DSML-notaties) empirisch met elkaar te kunnen vergelijken? (hoofdstuk 2);*
2. *Hoe wordt een methode opgezet die de intuïtiviteit van de ‘originele’ en ‘nieuwe’ PGA-notatie evalueert? (hoofdstuk 3);*
3. *In hoeverre wordt de ‘nieuwe’ PGA-notatie als intuïtiever beoordeeld in vergelijking met de ‘originele’ notatie? (hoofdstuk 4).*

1.5. Motivatie / relevantie

1.5.1. Wetenschappelijke relevantie

Het onderzoek draagt bij aan de body of knowledge over het empirisch evalueren van de intuïtiviteit van DSML-notaties. De eerste bijdrage is de opzet van een methode die de intuïtiviteit van verschillende versies van een DSML-notatie empirisch met elkaar vergelijkt. De tweede bijdrage is gelegen in het feit dat het onderzoek wordt gedaan in een praktische bedrijfscontext met daadwerkelijke eindgebruikers van de DSML – in tegenstelling tot vorig onderzoek. De derde bijdrage is het verkrijgen van inzicht in hoezeer de evaluatietechniek van Bork et al. (2019) tot een verbeterde DSML-notatie heeft geleid.

1.5.2. Maatschappelijke relevantie

Volgens McKinsey (2017) reageren organisatiemedewerkers beter op user interfaces en producten die visueel mooi en intuïtief bruikbaar zijn. Ook Gartner (2019) en BiZZDesign (2018) koppelen intuïtiviteit aan een verbeterde klantervaring. Hiermee kan geconcludeerd worden dat intuïtiviteit een positief effect heeft op gebruiksvriendelijkheid en acceptatie van DSML’s. Met andere woorden: een betere evaluatietechniek zal leiden tot intuïtievere DSML’s, die daardoor gebruiksvriendelijker is en waardoor eindgebruikers meer welwillend zijn om de DSML te gebruiken.

1.6. Aanpak in hoofdlijnen

Hoofdstuk 1 bevat de introductie met informatie over de achtergrond, probleemstelling en aanpak van het onderzoek. Hoofdstuk 2 bespreekt bestaande methoden die relevant zijn om de intuïtiviteit van notaties vergelijkend te evalueren. Deze inzichten resulteren in hoofdstuk 3 in een methodische opzet voor de vervolgevaluatie van de intuïtiviteit van de PGA-modelleertaal en de keuze voor een praktische bedrijfscontext. Na het uitvoeren van de methode bij een tiental respondenten binnen de praktische bedrijfscontext volgt in hoofdstuk 4 de analyse van de intuïtiviteit van de PGA-notatie op basis van de verkregen resultaten. In hoofdstuk 5 worden de resultaten samengevat en volgt een reflectie op de sterke en zwakke punten van de nieuwe methode.

2. Theoretisch kader

2.1. Onderzoeksaanpak

Het theoretisch kader moet een antwoord geven op de deelonderzoeksvraag: *‘Welke bestaande methoden zijn relevant om de intuïtiviteit van conceptuele modellen (i.e. DSML-notaties) empirisch met elkaar te kunnen vergelijken?’*. Het doel van het theoretisch kader is om inzicht te geven in de stand van zaken van de wetenschap over dit onderwerp, door hierover een samenvatting te geven van de kennis en ideeën, alsook de voor- en nadelen hiervan. Deze kennis leidt tot een betere onderzoeksfocus en de mogelijkheid om gestructureerd gegevens te verzamelen, onderzoeken en analyseren (Saunders, Lewis, & Thornhill, 2019).

2.1.1. Zoekstrategieën

Bij het opzetten van het theoretisch kader is gebruik gemaakt van de building blocks en forward snowballing methodes (Saunders et al., 2019). De building blocks methode wordt gebruikt wanneer de onderzoeker nog geen literatuur ter beschikking heeft, waarbij zoektermen worden afgeleid vanuit de onderzoeksvraag waarmee in een database kan worden gezocht. Vervolgens wordt per paper beoordeeld in hoeverre deze relevant is voor het beantwoorden van de onderzoeksvraag, aan de hand van vooraf opgestelde criteria. Op de meest relevante papers is vervolgens forward snowballing toegepast. Hierbij vormt een relevante paper of standaardwerk het startpunt en wordt gezocht naar publicaties die hierop hebben voortgebouwd.

2.1.2. Building blocks methode

Zoektermen en query

Vanuit de deelonderzoeksvraag zijn de kernconcepten ‘conceptuele modellen’, ‘empirisch vergelijken’ en ‘intuïtiviteit’ afgeleid. Hiertoe is in relevante databases gezocht met de Engelstalige zoektermen ‘conceptual modeling’, ‘empirical evaluation’ en ‘intuitivity’. Hieruit voortvloeiend is uiteindelijk de volgende query gebruikt:

- ("conceptual modeling") AND ("empirical evaluation") AND (intuitiv*)

Zoekopties

In tabel 1 staan de ingestelde zoekopties die automatisch op de zoekresultaten worden toegepast.

Tabel 1: Automatisch op de zoekresultaten toegepaste zoekopties

Zoekoptie		Toelichting
Gebruikte database - Building blocks	OU- Universiteitsbibliotheek	De OU-bibliotheek bevat toonaangevende wetenschappelijke tijdschriften, bibliografieën en bestanden, die voor OU studenten beschikbaar zijn. Dit maakt de kans groot dat deze databank voor dit onderzoek relevante papers bevat.
Gebruikte database - Forward snowballing	Web of Science	Web of Science is een multidisciplinair zoekplatform voor wetenschappelijke artikelen met een hoge kwaliteit. Door de OU-database wordt het aantal citaties in Web of Science weergegeven, wat kan worden gezien als een kwaliteitscheck.
Publicatiedatum	Vanaf 01-01-2004	De reden hiervoor is het publicatiejaar van Gemino & Wand (2004) dat wordt gezien als een basiswerk over het empirisch vergelijken van conceptuele modellen. Dit maakt papers van 2004 en jonger relevant voor het beantwoorden van de onderzoeksvraag
Taal	Engels	Dit is de standaardtaal van het onderzoeksgebied
Resultaten uitsluiten	Krantenartikelen, book reviews	Voor het beantwoorden van de deelonderzoeksvraag zijn alleen wetenschappelijke papers van belang. Krantenartikelen en book reviews zijn dus niet relevant
Uw zoekopdracht verfijnen	Volledige tekst online	Hierdoor is van de query-resultaten de volledig tekst online te vinden

Inclusie-/exclusiecriteria

De relevantie van de papers is bepaald op basis van de inclusie-/exclusiecriteria in tabel 2, die handmatig werden toegepast op de zoekresultaten.

Tabel 2: Handmatig op de zoekresultaten toegepaste inclusie-/exclusiecriteria

Inclusie-/exclusie criterium		Toelichting
Aantal referenties in Web of Science	Hoog aantal	Papers worden als relevanter beschouwd wanneer er vaak naar wordt gerefereerd. Bij het prioriteren van papers worden papers met een hogere aantal referenties als relevanter beschouwd dan onderzoeken met een lager aantal
Jaar van publicatie	Recent	Bij het prioriteren van papers wordt een paper die recenter is als meer relevant beschouwd dan een paper die ouder is. De reden hiervoor is gelegen in de aanname dat jongere papers meer recente kennis en ideeën bevatten over het onderzoeksveld dan oudere papers
Soort paper	Methodologie of individueel experiment	Papers worden alleen als relevant gezien als deze gaan over methodologie of over een individueel experiment die te maken heeft met het onderwerp van de deelonderzoeksvraag
Trefwoorden in de tekst	Link met intuïtiviteit	Een paper is meer relevant wanneer in de tekst een koppeling wordt gemaakt tussen het onderzoek en het trefwoord 'intuitivity'. Indien dit niet het geval is, dan moet de paper minimaal één van onderstaande typen vragen bevatten
Gebruikte taaktypen	Recall, comprehension, problem-solving, cloze	Een paper is relevant wanneer deze één van de hiernaast genoemde taaktypen bevat. Dit zijn immers de door Gemino and Wand (2004) voorgestelde taaktypen voor het empirisch evalueren van conceptuele modellen. Bij het prioriteren van papers geldt: hoe meer typen vragen de paper bevat, hoe relevanter deze is
Soort respondenten	Business users	Papers waarin de respondenten business users zijn, worden als meer relevant beschouwd dan studenten. De eindgebruikers van de PGA-notatie zijn immers business users
Stopcriterium: - Building blocks	Er is gestopt met zoeken naar relevante artikelen na de vierde opeenvolgende paper die niet voldoet aan bovengenoemde criteria	Het stopcriterium is vastgesteld op basis van de zoekervaringen naar relevante papers. Toen is namelijk gebleken dat wanneer vier opeenvolgende papers niet aan de inclusie-/exclusie criteria voldoen, dat de hierop volgende artikelen hier dan ook niet aan voldoen.
Stopcriterium: - Forward snowballing	Alle papers zijn beoordeeld	De keuze om alle papers te beoordelen is gelegen in het feit dat sortering op de Web of Science webpagina plaatsvindt op het aantal referenties in Web of Science. Hierdoor kan niet zondermeer geconcludeerd worden dat papers die verder onderaan in de lijst staan per definitie minder relevant zijn.

2.1.3. Forward snowballing methode

Op de meest relevante papers (i.e. die voldoen aan criterium over link met intuïtiviteit) die zijn voortgekomen uit het toepassen van de building blocks methode is forward snowballing toegepast. Per artikel is gekeken welke onderzoeken naar het oorspronkelijke artikel hebben verwezen binnen Web of Science, gesorteerd op afnemend aantal referenties in Web of Science. Deze artikelen zijn vervolgens op relevantie beoordeeld op basis van de inclusie-/exclusiecriteria in tabel 2.

2.2. Uitvoering

Onderstaande beschrijving bevat de resultaten, alsook de referenties van de zes uiteindelijk relevante papers. De gedetailleerde analyse is te vinden in bijlage 2.

De selectie is tot stand gekomen door onderscheid te maken in het al dan niet voldoen aan het criterium over de link met intuïtiviteit, omdat dit als meest relevant worden gezien om de deelonderzoeksvraag te kunnen beantwoorden. Bij vijf van de uiteindelijke zes papers is deze link aanwezig. Drie papers zijn naar voren gekomen uit de query, twee papers uit forward snowballing op de resultaten van de query. Van de papers die voor het onderwerp relevant zijn - zonder link met intuïtiviteit - is de paper van Aranda, Ernst, Horkoff, and Easterbrook (2007) gekozen, omdat de titel trefwoorden (i.e. empirical evaluation en comprehensibility) bevat die het meest relevant zijn voor het beantwoorden van de onderzoeksvraag. Empirisch evalueren (i.e. empirical evaluation) is immers onderdeel van de deelonderzoeksvraag en begrijpelijkheid (i.e. comprehensibility) is nauw verwant aan intuïtiviteit (i.e. intuïtieve begrijpelijkheid).

2.2.1. Resultaten Building Blocks methode

Tabel 3 bevat de resultaten en tabel 4 de relevante referenties van de query: ("conceptual modeling") AND ("empirical evaluation") AND (intuïtiv*).

Tabel 3: Resultaten building blocks query

Kenmerk	Aantal
Aantal zoekresultaten	40
Aantal papers op criteria beoordeeld	17
Aantal relevante papers: link met intuïtiviteit	3
Aantal relevante papers: geen link met intuïtiviteit	4

Tabel 4: Relevante referenties building blocks query

Referentie	Titel
Moody, Heymans, and Matulevičius (2010)	Visual syntax does matter: improving the cognitive effectiveness of the i* visual notation
Figl, Mendling, and Strembeck (2013)	The influence of notational deficiencies on process model comprehension
Figl (2017)	Comprehension of procedural visual business process models
Aranda et al. (2007)	A framework for empirical evaluation of model comprehensibility

2.2.2. Resultaten Forward snowballing methode

Op de drie papers die een link hebben met intuïtiviteit (Figl, 2017; Figl et al., 2013; Moody et al., 2010) is forward snowballing toegepast. Dit heeft voor alleen Figl et al. (2013) geleid tot twee extra relevante papers die eveneens een link hebben met intuïtiviteit, zie tabel 5 en 6 voor respectievelijk de resultaten en relevante referenties.

Tabel 5: Resultaten forward snowballing op Figl et al. (2013)

Kenmerk	Resultaat
Aantal zoekresultaten	18
Aantal papers op criteria beoordeeld	18
Aantal relevante papers: link met intuïtiviteit	3 (waarvan er één al voorkwam in de building blocks query)
Aantal relevante papers: geen link met intuïtiviteit	4

Tabel 6: Relevante referenties forward snowballing op Figl et al. (2013)

Referentie	Titel
Figl and Recker (2016)	Exploring cognitive style and task-specific preferences for process representations
Jošt, Huber, Heričko, and Polančič (2016)	An empirical investigation of intuitive understandability of process diagrams

2.3. Resultaten en conclusies

Het theoretisch kader is opgebouwd rondom twee methodologische papers (Aranda et al., 2007; Figl, 2017), één evaluatie van een notatie door middel van desk research (Moody et al., 2010) en drie vergelijkende evaluaties van notaties door middel van een experiment (Figl et al., 2013; Figl & Recker, 2016; Jošt et al., 2016). Allereerst volgt een beknopte samenvatting over de methodologische aspecten: (i) maatstaven; (ii) experimenteel design; (iii) experimentele taken; (iv) participanten; (v) procedure; en (vi) raamwerk. Een gedetailleerde analyse van deze aspecten per paper is te vinden in bijlage 3. Voor de (i) maatstaven wordt beschreven welke afhankelijke en onafhankelijke variabelen in het onderzoek worden gebruikt. Voor de experimentele papers wordt bij het (ii) experimenteel design duidelijk gemaakt of in het experiment gebruik gemaakt wordt van een within-subjects of between-subjects design. Volgens Saunders et al. (2019) wordt het experiment in geval van een within-subjects design uitgevoerd binnen één groep, waarbij alle participanten worden blootgesteld aan de geplande interventie. Bij een between-subjects design wordt gebruik gemaakt van zowel een experimentele groep als een controlegroep, waarbij alleen de experimentele groep de interventie ondergaat. Door deze opzet zijn bij dit type design de onderzoeksresultaten tussen de experimentele en controlegroep met elkaar te vergelijken, waarbij het voor de hand ligt dat eventuele verschillen verklaard kunnen worden door de interventie. Voor wat betreft de (iii) experimentele taken wordt beschreven van welk type taken gebruik wordt gemaakt. Daarbij zal specifiek worden gekeken naar de door Gemino and Wand (2004) voorgestelde taaktypen voor het empirisch evalueren van conceptuele modellen: begripstaken (i.e. comprehension), herinneringstaken (i.e. recall), semantische herinneringstaken (i.e. cloze test) en probleemoplossingstaken (i.e. problem-solving). Begripstaken bestaan doorgaans uit een reeks vragen over notatie-elementen waarbij de notatie wordt getoond. Bij herinneringstaken worden vragen gesteld over de notatie, nadat deze is weggenomen. In het geval van semantische herinneringstaken wordt eveneens de notatie weggenomen, waarna de participant

wordt gevraagd om in een tekst de lege plekken in te vullen. Probleemoplossingstaken zijn begripstaken die gericht zijn op domeinbegrip. In geval van de (iv) participanten wordt specifiek gekeken naar de in de onderzoeken gebruikte participanten: studenten of business users. Daarnaast wordt bij experimentele papers het aspect (v) procedure beschreven en bij methodologische en desk research papers het (vi) raamwerk (i.e. framework) dat in het onderzoek wordt gebruikt.

Vervolgens worden per aspect conclusies getrokken, waarmee een antwoord wordt gegeven op deelvraag: *‘Welke bestaande methoden zijn relevant om de intuïtiviteit van conceptuele modellen (i.e. DSML-notaties) empirisch met elkaar te kunnen vergelijken?’*.

2.3.1. Maatstaven

Objectieve Maatstaven Intuïtiviteit

Tabel 7 bevat een overzicht van de objectieve maatstaven die in de papers gebruikt worden om de afhankelijke variabele intuïtiviteit (i.e. begrijpelijkheid of intuïtieve begrijpelijkheid) te meten. Opvallend is dat er sprake lijkt te zijn van twee veel voorkomende maatstaven: interpretatieve effectiviteit en interpretatieve efficiëntie, die ook terug te vinden zijn in het state of the art artikel van Figl (2017) en eveneens worden gebruikt door Jošt et al. (2016), die expliciet intuïtieve begrijpelijkheid meten. Daarnaast maken Jošt et al. (2016) ook gebruik van de maatstaf interpretatieve doeltreffendheid, een ratio van interpretatieve effectiviteit en interpretatieve efficiëntie, die afkomstig is van een reeds door Rolon et al. (2009) gevalideerde set variabelen voor het evalueren van de begrijpelijkheid van een procesdiagram.

Tabel 7: Overzicht objectieve maatstaven intuïtiviteit

Auteur	Afhankelijke variabele	Maatstaven
Aranda et al. (2007)	Begrijpelijkheid	Interpretatieve Effectiviteit
Figl (2017)	Begrijpelijkheid	Interpretatieve Effectiviteit
Moody (2010)	Begrijpelijkheid	Interpretatieve Effectiviteit
Figl et al. (2013)	Begrijpelijkheid	Interpretatieve Effectiviteit
Jost et al. (2016)	Intuïtieve begrijpelijkheid	Interpretatieve Effectiviteit
Aranda et al. (2007)	Begrijpelijkheid	Interpretatieve Efficiëntie
Figl (2017)	Begrijpelijkheid	Interpretatieve Efficiëntie
Moody (2010)	Begrijpelijkheid	Interpretatieve Efficiëntie
Figl et al. (2013)	Begrijpelijkheid	Interpretatieve Efficiëntie
Jost et al. (2016)	Intuïtieve begrijpelijkheid	Interpretatieve Efficiëntie
Jost et al. (2016)	Intuïtieve begrijpelijkheid	Interpretatieve Doeltreffendheid

Subjectieve Maatstaven Intuïtiviteit

Tabel 8 bevat een overzicht van de subjectieve maatstaven waarmee de afhankelijke variabele intuïtiviteit (i.e. begrijpelijkheid of intuïtieve begrijpelijkheid) wordt gemeten. Opvallend is dat vaak gebruik wordt gemaakt van variabele waargenomen gebruiksgemak. De papers van Figl (2017) en Figl and Recker (2016) voegen daar ook de maatstaf representatievoorkeur aan toe. Tot slot gebruiken Aranda et al. (2007) daarnaast ook de maatstaf subjectieve interpretatieve effectiviteit.

Tabel 8: Overzicht subjectieve maatstaven intuïtiviteit

Auteur	Afhankelijke variabele	Maatstaven
Figl (2017)	Begrijpelijkheid	Waargenomen gebruiksgemak
Moody (2010)	Begrijpelijkheid	Waargenomen gebruiksgemak
Figl et al. (2013)	Begrijpelijkheid	Waargenomen gebruiksgemak
Aranda et al. (2007)	Begrijpelijkheid	Waargenomen gebruiksgemak
Figl (2017)	Begrijpelijkheid	Representatievoorkeur
Figl & Recker (2016)	Begrijpelijkheid	Representatievoorkeur
Aranda et al. (2007)	Begrijpelijkheid	Subjectieve interpretatieve effectiviteit

Onafhankelijke Variabelen

Tabel 9 bevat een overzicht van onafhankelijke variabelen. Opvallend is dat alle variabelen zijn te matchen met de variabelen die door Figl (2017) zijn gedefinieerd en dat vooral gebruik is gemaakt van variabelen omtrent notatie, gebruikerseigenschappen, modeleigenschappen en secundaire notatie. Ook Jošt et al. (2016) zien notatie (i.e. primaire notatie) en modeleigenschappen (i.e. diagramcomplexiteit) als factoren die van invloed zijn op intuïtiviteit. Moody et al. (2010) maken in hun onderzoek een expliciete koppeling tussen intuïtiviteit en de notatie variabele semantic transparency. Daarbij zou perceptual discriminability een positief interactie-effect, en cognitive fit een positief of negatief interactie-effect hebben op semantic transparency. In het artikel wordt echter niet duidelijk gemaakt waar dat precies van afhangt.

Tabel 9: Overzicht onafhankelijke variabelen die van invloed zijn op intuïtiviteit

Cluster	Auteur	Onafhankelijke variabele
Notatie	Moody (2010)	Cognitive fit
Notatie	Moody (2010)	Cognitive integration
Notatie	Moody (2010)	Complexity management
Notatie	Moody (2010)	Graphic economy
Notatie	Figl (2017)	Notatie-eigenschappen
Notatie	Figl et al. (2013)	Perceptual discriminability
Notatie	Moody (2010)	Perceptual discriminability
Notatie	Jost et al. (2016)	Primaire notatie
Notatie	Figl (2017)	Primaire notatie
Notatie	Figl (2017)	Representatie paradigma
Notatie	Moody (2010)	Semantic transparency
Notatie	Figl et al. (2013)	Semiotic clarity
Notatie	Moody (2010)	Semiotic clarity
Notatie	Moody (2010)	Visual expressiveness
Gebruikerseigenschappen	Aranda et al. (2007)	Domeinbegrip
Gebruikerseigenschappen	Figl (2017)	Domeinbegrip
Gebruikerseigenschappen	Figl & Recker (2016)	Cognitieve stijl
Gebruikerseigenschappen	Aranda et al. (2007)	Modelleerkennis
Gebruikerseigenschappen	Figl (2017)	Modelleerkennis
Gebruikerseigenschappen	Figl (2017)	Modelleerervaring
Gebruikerseigenschappen	Figl & Recker (2016)	Modelleerervaring
Gebruikerseigenschappen	Figl (2017)	Opleidingsniveau
Modeleigenschappen	Figl (2017)	Connecties
Modeleigenschappen	Jost et al. (2016)	Diagramcomplexiteit
Modeleigenschappen	Aranda et al. (2007)	Domeingrootte
Modeleigenschappen	Figl (2017)	Gateway-samenspel
Modeleigenschappen	Figl (2017)	Maat
Modeleigenschappen	Figl (2017)	Structuur
Modeleigenschappen	Figl (2017)	Syntax regels
Secundaire notatie	Figl (2017)	Decompositieregels
Secundaire notatie	Moody (2010)	Dual coding
Secundaire notatie	Figl (2017)	Gestalt theorie
Secundaire notatie	Figl (2017)	Layout
Taaksoort	Aranda et al. (2007)	Taaksoort
Taaksoort	Figl & Recker (2016)	Taaksoort
Taaksoort	Figl (2017)	Taaksoort
Naamgeving	Figl (2017)	Label design
Naamgeving	Figl (2017)	Naamgevingsconventies
Presentatiemedium	Figl (2017)	Presentatiemedium

2.3.2. Experimenteel Design

Figl (2017) geeft aan dat 50% van de onderzoeken gebruik maakt van een between-subjects design, 42% van een within-subjects design en 8% van een gecombineerd design. De keuze voor het experimenteel design is afhankelijk van de typen variabelen die onderzocht worden. Bij between-subjects worden vaak de variabelen notatie, secundaire notatie en gebruikerseigenschappen gebruikt, bij within-subjects de variabelen modeleigenschappen en taaksoort. Tabel 10 bevat een overzicht van de experimentele designs waarvan in de onderzoeken gebruik is gemaakt.

Tabel 10: Overzicht experimentele designs

Auteur	Experimenteel design
Figl et al. (2013)	Between-subjects
Jost et al. (2016)	Between-subjects
Figl & Recker (2016)	Within-subjects
Aranda et al. (2007)	N.v.t. het betreft een methodologische paper
Figl (2017)	N.v.t. het betreft een methodologische paper
Moody (2010)	N.v.t. het betreft een notatie evaluatie door middel van desk research

2.3.3. Experimentele Taken

Volgens Figl (2017) zijn er twee begripstypen: surface-level understanding (i.e. de mate waarin een persoon een notatie begrijpt) en deep-level understanding (i.e. mate waarin een persoon de notatie toe kan passen). In het onderzoek wordt gesteld dat onderzoekers zich het beste kunnen richten op het meten van surface-level understanding met begripstaken (i.e. comprehension tasks), omdat dit de meest gebruikte meetmethode is en onderzoeksresultaten daarom makkelijker met elkaar vergeleken kunnen worden. Daarnaast zijn metingen meer direct te relateren aan concrete modellen dan metingen via herinneringstaken (i.e. recall tasks) of probleemoplossingstaken (i.e. problem-solving tasks).

Verder stelt Figl (2017) dat surface-level understanding overwegend wordt gemeten met multiple-choice begripstaken. Bij deze taken zou de onderzoeker informatie moeten verstrekken die overeenkomt met de informatiebehoefte van een gebruiker bij een real-life gebruik van het model. Naast multiple-choice begripstaken taken spreekt Figl (2017) ook over nieuwe meetmethoden uit de hoek van de neurowetenschappen. Een simpele neurofysiologische tool die reeds succesvol is ingezet om begripelijkheid te meten is eye-tracking, waarbij de oogfixatie op een relevante regio van de notatie voor 70% de begripstaakscores kan voorspellen. Een andere opkomende methode is het meten van hersenactiviteit door gebruik te maken van neuro-imaging tools, die de potentie hebben om begripelijkheid objectiever en fijnmaziger te meten dan met de multiple-choice begripstaken.

Tabel 11 bevat de experimentele taken waarvan in de onderzoeken gebruik is gemaakt. Hierin is te zien dat Figl et al. (2013) en Jošt et al. (2016) gebruik maken van begripstaken voor het vergelijkend evalueren van respectievelijk begripelijkheid en intuïtieve begripelijkheid van notaties, en dat Figl and Recker (2016) representatievoorkeur taken gebruiken om procesrepresentaties vergelijkend te evalueren.

Tabel 11: Overzicht experimentele taken

Auteur	Experimentele taken
Figl et al. (2013)	Begripstaken
Jost et al. (2016)	Begripstaken
Figl & Recker (2016)	Representatievoorkeur taken
Aranda et al. (2007)	N.v.t. het betreft een methodologische paper
Figl (2017)	N.v.t. het betreft een methodologische paper
Moody (2010)	N.v.t. het betreft een notatie evaluatie door middel van desk research

2.3.4. Participanten

Figl (2017) geeft aan dat participanten in 62% van de onderzoeken bestaan uit studenten, 13% uit domein experts, 4% uit proces model experts vanuit de academische wereld alsook uit de praktijk en voor 21% uit gemengde groepen. Tabel 12 bevat een overzicht van de participanten waarvan gebruik is gemaakt in de onderzoeken.

Tabel 12: Overzicht participanten

Auteur	Participanten
Figl et al. (2013)	Studenten zonder voorkennis over procesmodellieren
Jost et al. (2016)	Studenten zonder voorkennis over onderzochte notaties
Figl & Recker (2016)	Business school studenten
Aranda et al. (2007)	N.v.t. het betreft een methodologische paper
Figl (2017)	N.v.t. het betreft een methodologische paper
Moody (2010)	N.v.t. het betreft een notatie evaluatie door middel van desk research

2.3.5. Procedures

Tabel 13 bevat een overzicht met procedurestappen die in de onderzoeken met een experiment zijn uitgevoerd. Hierin is de volgende fasering zichtbaar: (i) inleiding; (ii) uitvragen gebruikerseigenschappen en modelleerervaring; (iii) verstrekken relevante informatie over de notatie; (iv) presenteren en uitvoeren van experimentele taken; en (v) afsluiting.

Tabel 13: Overzicht experimentele procedurestappen

Cluster	Auteur	Procedurestappen experiment
Inleiding	Jost et al. (2016)	Inleidende toelichting
Inleiding	Jost et al. (2016)	Verstrekken paper-based instructies
Inleiding	Figl & Recker (2016)	Tonen informatie met toestemmingformulier en toelichting
Gebruikerseigenschappen	Figl et al. (2013)	Vragen over demografische gegevens, academische kwalificaties en modelleerervaring
Gebruikerseigenschappen	Jost et al. (2016)	Beantwoorden vragen over descriptive statistics (participant)
Gebruikerseigenschappen	Figl & Recker (2016)	Vullen demografische gegevens
Modelleerervaring	Figl & Recker (2016)	Vullen onderdeel over ervaring met conceptuele modellen
Modelleerervaring	Figl et al. (2013)	Test over algemene kennis over procesmodelleren
Notatie informatie	Figl et al. (2013)	Tutorial met relevante informatie over de notatie
Begripstaken	Figl et al. (2013)	Verschillende notaties met bijbehorende begripstaken
Begripstaken	Jost et al. (2016)	Oplossen van vragen over proces (participant)
Begripstaken	Figl & Recker (2016)	Invullen OSIVQ voor cognitive style
Begripstaken	Figl & Recker (2016)	Tonen verschillende notaties
Begripstaken	Figl & Recker (2016)	Stellen vragen over taakspecifieke notatievoorkeur
Afsluiting	Jost et al. (2016)	Inleveren survey (participant)

2.3.6. Raamwerk

De paper van Aranda et al. (2007) biedt een theoretisch en methodologisch gegrond raamwerk voor het empirisch evalueren van de begrijpelijkheid van procesmodellen. Het raamwerk bestaat uit een zevental stappen: (i) het selecteren van de modelleernotatie; (ii) het beschrijven van de theorie die aan de notatie ten grondslag ligt; (iii) het doorvertalen hiervan naar claims over de begrijpelijkheid van de notatie; (iv) het kiezen van een baseline waarmee de notatie wordt vergeleken; (v) het omzetten van begrijpelijkheidclaims naar te testen hypothesen; (vi) het verrijken van hypothesen met inzichten uit begrijpelijkheidstheorieën (i.e. external cognition en cognitive dimensions framework); en (vii) het ontwerpen en uitvoeren van het experiment. Belangrijk in dit onderdeel is om te kiezen voor specifieke - door modelleerexperts gemaakte - modellen in een voor de notatie en participanten natuurlijk domein. In de paper wordt niet concreet gemaakt hoe stap (vii) er precies uitziet.

Moody et al. (2010) doorlopen bij de notatieanalyse de volgende stappen: (i) definiëren van een principe (i.e. onafhankelijke variabele); (ii) beschrijven resultaten van de evaluatie; (iii) presenteren aanbevelingen voor verbetering; en (iv) wanneer van toepassing, analyseren interactie met andere principes.

2.3.7. Conclusie

Tabel 14 bevat een overzicht met methodes die het meest relevant zijn om de intuïtiviteit van notaties vergelijkend empirisch te evalueren.

Objectieve Maatstaven Intuïtiviteit

De meest voorkomende en eveneens door Jošt et al. (2016) gebruikte objectieve maatstaven voor de afhankelijke variabele intuïtieve begrijpelijkheid zijn: interpretatieve effectiviteit en interpretatieve efficiëntie.

Subjectieve Maatstaven Intuïtiviteit

De meest gebruikte subjectieve maatstaven zijn waargenomen gebruiksgemak en representatievoorkeur.

Onafhankelijke variabelen

Tabel 14 bevat de meest voorkomende categorieën van onafhankelijke variabelen: (i) notatie; (ii) secundaire notatie (iii); gebruikerseigenschappen; en (iv) modeleigenschappen. Bij de (i) notatie kan specifiek worden gekeken naar de door Jošt et al. (2016) gebruikte primaire notatie, alsook naar semantic transparency en perceptual discriminability die door Moody et al. (2010) in verband worden gebracht met intuïtiviteit. Bij de (ii) secundaire notatie wordt in de papers geen specifieke richting gegeven waarnaar gekeken zou moeten worden. Voor wat betreft de (iii) gebruikerseigenschappen kunnen alle door Figl (2017) aangegeven factoren worden meegenomen: domeinbegrip, modelleerkennis, modelleerervaring en opleidingsniveau. In geval van (iv) modeleigenschappen kan worden gekeken naar de door Jošt et al. (2016) gebruikte diagramcomplexiteit.

Experimenteel design

Aangezien in de papers overwegend experimentele methoden worden gebruikt voor het empirisch evalueren van de intuïtieve begrijpelijkheid van notaties kan het experiment worden gezien als de meest geschikte methode om hypothesen hierover te testen. Als experimenteel design lijkt de between-subjects het meest relevant, omdat deze het meest geschikt zou zijn voor het onderzoeken van de onafhankelijke variabelen (i) notatie, (ii) secundaire notatie en (iii) gebruikerseigenschappen.

Experimentele taken

Aangezien de afhankelijke variabele intuïtieve begrijpelijkheid wordt gedefinieerd als “*het gemak waarmee een notatie direct kan worden begrepen door gebruikers, zonder enige voorkennis of training*” (Jošt et al., 2016, p. 91) zou het experiment uitgevoerd kunnen worden met surface-level multiple choice begripstaken. Wanneer specifiek de voorkeur van een notatie boven een andere notatie wordt uitgevraagd, dan kan ook gebruik gemaakt worden van de representatievoorkeur taken van Figl and Recker (2016) om notaties vergelijkend te evalueren.

Participanten

Gezien de definitie van intuïtieve begrijpelijkheid zouden participanten zonder enige voorkennis over of training in de notatie (Jošt et al., 2016) in een experiment de voorkeur hebben.

Procedure

Tijdens het experiment kan de procedure worden gevolgd met de meest voorkomende vijf stappen, die in tabel 14 staan beschreven.

Raamwerk

De eveneens in tabel 14 beschreven zeven stappen uit het onderzoekraamwerk van Aranda et al. (2007) is te gebruiken als overkoepelend raamwerk voor het empirisch evalueren van de intuïtiviteit van notaties.

Tabel 14: Totaaloverzicht relevante experimentele methodes

Methodologisch aspect	Relevante experimentele methodes
Afhankelijke variabele - objectieve maatstaven	- Interpretatieve effectiviteit - Interpretatieve efficiëntie
Afhankelijke variabele - subjectieve maatstaven	- Representatievoorkeur - Waargenomen gebruiksgemak
Onafhankelijke variabele	- Notatie (i.e. primaire notatie, semantic transparency, perceptual discriminability) - Secundaire notatie - Gebruikerseigenschappen (i.e. domeinbegrip, modelleerkennis, modelleerervaring, opleidingsniveau) - Modeleigenschappen (i.e. diagramcomplexiteit)
Experimenteel design	Between-subjects
Experimentele taken	Surface-level multiple-choice begripstaken en eventueel Representatievoorkeur taken
Participanten	Gebruikers zonder voorkennis over of training in de onderzochte notatie
Procedure	Procedurestappen: (i) Inleiding (ii) Uitvragen gebruikerseigenschappen en modelleerervaring (iii) Verstrekken relevante informatie over de notatie (iv) Presenteren en uitvoeren van experimentele taken (v) Afsluiting
Raamwerk	Zeven stappen in het raamwerk van Aranda et al. (2007) voor het empirisch evalueren van de begripelijkheid van notaties: (i) Selecteren modelleernotatie (ii) Beschrijven aan notatie ten grondslag liggende theorie (iii) Doorvertalen naar claims over begripelijkheid notatie (iv) Kiezen baseline waarmee notatie wordt vergeleken (v) Omzetten begripelijkheidclaims naar hypothesen (vi) Verrijken hypothesen met inzichten uit begripelijkheidstheorieën (vii) Ontwerpen en uitvoeren experiment

2.4. Doel van het vervolgonderzoek

In voorgaand hoofdstuk is vastgesteld welke methoden er bestaan om de intuïtiviteit van conceptuele modellen (i.e. DSML-notaties) empirisch met elkaar te vergelijken en welke combinatie van methoden hierbij het meest relevant zou zijn. Het doel van het vervolgonderzoek is om op basis van deze inzichten een vervollexperiment op te zetten die de intuïtiviteit van de 'originele' en verbeterde, 'nieuwe' PGA-notatie evalueert in een praktische bedrijfscontext. Vervolgens zal het experiment worden uitgevoerd bij een tiental respondenten en vindt op basis van de experimentele onderzoeksresultaten een analyse plaats op de intuïtiviteit van de PGA-notaties. Tot slot zal een conclusie worden getrokken in hoeverre de 'nieuwe' PGA-notatie intuïtiever is dan de 'originele' en zal daarnaast ook een uitspraak worden gedaan over de mate waarin de evaluatietechniek van Bork et al. (2019) heeft geleid tot significante verbeteringen in de intuïtiviteit van de PGA-notatie voor eindgebruikers in een praktische bedrijfscontext. Eveneens zal een besluit worden genomen over eventueel benodigde aanpassingen aan de experimentele opzet van dit onderzoek.

3. Methodologie

3.1. Conceptueel ontwerp: keuze van onderzoeksmethode(n)

In dit deel van het onderzoek wordt een antwoord gegeven op deelonderzoeksvraag ‘*Hoe wordt een methode opgezet die de intuïtiviteit van de ‘originele’ en ‘nieuwe’ PGA-notatie evalueert?*’. Hiertoe in dit hoofdstuk een methode opgezet die de intuïtiviteit van de ‘originele’ en de op basis van de evaluatietechniek verbeterde ‘nieuwe’ PGA-notatie evalueert. Dit wordt gedaan door middel van een vergelijkende evaluatie van de verschillende versies van de PGA-notatie binnen een praktische bedrijfscontext. In de probleemstelling is immers vastgesteld dat de analyse en het ontwerp van de verbeterde notatie door middel van de evaluatietechniek van Bork et al. (2019) deels afhankelijk is van menselijke creativiteit. Hierdoor is niet met zekerheid te zeggen of de techniek tot significante verbeteringen heeft geleid qua intuïtiviteit van een DSML-notatie voor eindgebruikers binnen een praktische bedrijfscontext.

Om het doel te realiseren is informatie nodig uit de wetenschappelijke literatuur over relevante methoden om de intuïtiviteit van notaties vergelijkend te evalueren, die is samengevat in het theoretisch kader. Daarnaast is een gedetailleerde onderzoeksmethode nodig voor de vergelijkende evaluatie van de intuïtiviteit van de PGA-modelleertaal, die in het methodologische hoofdstuk wordt uitgewerkt. Tot slot zijn vergelijkende evaluatieresultaten nodig binnen de praktische bedrijfscontext. Een deductieve onderzoekaankpak is in staat om deze informatie te leveren. Het startpunt van deze gestructureerde aanpak is volgens Saunders et al. (2019) namelijk methodologische literatuur, waarbij ervanuit wordt gegaan dat deze theorie ook van toepassing is op de specifieke onderzoeksituatie. Om dit te testen worden op basis van de literatuur variabelen geoperationaliseerd en omgezet naar hypothesen. Na de dataverzameling wordt vervolgens geanalyseerd in hoeverre de data de hypothesen ondersteunen en tot op welke hoogte de voorgestelde theorie juist is en gegeneraliseerd kan worden, of moet worden aangepast.

Volgens het theoretisch kader is een experiment de meest geschikte methode om hypothesen over het evalueren van de intuïtieve begrijpelijkheid van notaties te testen. Saunders et al. (2019) beschrijven het doel van het experiment als “*het bestuderen van de kans dat een verandering van een onafhankelijke variabele, een verandering in een afhankelijke variabele tot gevolg heeft*” (p. 803). Hiertoe worden nulhypothesen (i.e. er is géén relatie tussen variabelen) en alternatieve hypothesen (i.e. er is een relatie tussen variabelen) geformuleerd.

Overigens wordt in twee separate onderzoeken, bij twee andere casusorganisaties, ditzelfde experiment tegelijkertijd uitgevoerd. Dit heeft als consequentie dat er in dit hoofdstuk experimentele taken worden gebruikt (i.e. notatieassociatie- en herinneringstaken) die in het theoretisch kader van dit onderzoek niet naar voren zijn gekomen, maar in het theoretisch kader van de twee separate onderzoeken wel.

3.2. Technisch ontwerp: uitwerking van de methode

3.2.1. Variabelen & maatstaven

In dit onderzoek wordt de afhankelijke variabele intuïtieve begrijpelijkheid gedefinieerd als *“het gemak waarmee gebruikers het diagram direct kunnen begrijpen, zonder enige voorkennis of training”* (Jošt et al., 2016, p. 91). Deze variabele wordt gemeten met objectieve maatstaven (i) interpretatieve effectiviteit; (ii) interpretatieve efficiëntie; en de subjectieve maatstaf (iii) representatievoorkeur (Figl, 2017). De (i) interpretatieve effectiviteit wordt gemeten door het aantal correcte antwoorden op de experimentele taken. De (ii) interpretatieve efficiëntie wordt vastgesteld door de tijd (in seconden) te meten die de participant nodig heeft om de experimentele taken af te ronden. Bij de (iii) representatievoorkeur wordt de participant gevraagd naar welk versie van de PGA-notatie de intuïtieve voorkeur uitgaat. Hierbij wordt gebruik gemaakt van de reeds door Figl and Recker (2016) toegepaste schaal van Cook, Heath, Thompson, and Thompson (2001), die loopt van 0 (e.g. voorkeur gaat naar de ‘originele’ notatie) tot 100 (e.g. voorkeur gaat naar de ‘nieuwe’ notatie), waarbij 50 betekent dat er geen specifieke voorkeur wordt aangegeven. De onafhankelijke variabele in het onderzoek is de notatie (i.e. ‘nieuwe’ en ‘originele’ PGA-notatie), die door Bork et al. (2019) wordt omschreven als de grafische representatie van een conceptueel model. Het onderzoeksdoel is immers om twee versies van de PGA-notatie vergelijkend te evalueren. Die vergelijking zal specifiek worden gemaakt voor de door (Roelens & Bork, 2020) verbeterde, en in bijlage 4 weergegeven, elementen: value proposition; internal goal; competence; customer goal; en de combinatie van importance en value stream.

3.2.2. Experimentele taken

Het experiment start met een instemmingsformulier voor deelname aan het onderzoek en een reeks (i) demografische vragen. Hierop volgen de experimentele taken: (ii) notatieassociatie taken; (iii) begripstaken; (iv) herinneringstaken; en (v) representatievoorkeur taken; zie bijlage 5 voor de tijdens het experiment gebruikte survey in Limesurvey. Bij de (i) demografische vragen wordt de participant gevraagd naar persoonsgegevens (i.e. geslacht, geboortjaar, opleidingsniveau, werksector) en modelleerervaring (i.e. modelleerkennis, modelleer expertise, modelleerervaring). Hiermee kan worden onderzocht tot op welke hoogte participanten voorkennis en/of training hebben gehad in het modelleren van conceptuele modellen en in het specifieke domein. Bij de (ii) notatieassociatie taken zal de participant - specifiek voor het meten van semantic transparency - worden gevraagd om na het zien van een notatie drie intuïtieve associaties vast te leggen die bij de participant opkomen, uiteraard zonder vooraf hints te hebben geven over de naam of de semantiek (Bork et al., 2018). In dit onderdeel van het experiment moet erop worden gelet dat beide experimentele groepen voorzien worden van precies dezelfde informatie, zodat uitgesloten kan worden dat dit een invloedfactor is in de onderzoeksresultaten. De (iii) begripstaken bestaan uit een set vragen over een gegeven model (Figl, 2017) met bijgevoegd precies de informatie die nodig is om het begrip bij een min of meer realistisch gebruik te simuleren, zie bijlage 6. Het gegeven model beschrijft met behulp van de PGA-notatie hoezeer de processen van een organisatie op de doelen zijn uitgelijnd. Aangezien in dit onderzoek wordt gewerkt met surface-level multiple-choice begripstaken (Figl, 2017) met vragen over de semantiek van de notatie (Jošt et al., 2016) is gekozen voor een organisatie waarvan wordt verwacht dat participanten hiermee bekend zijn: een online marktplaats voor het aanbieden van eten door restaurantpartners en het bestellen ervan door klanten. Hierdoor kan worden uitgesloten dat een gebrek aan specifieke werksectorkennis een invloedfactor is voor het fout beantwoorden van de

begripstaken. Bij de (iv) herinneringstaken wordt het PGA-notatie element en de onderliggende betekenis (i.e. korte definitie) getoond en vervolgens weer weggelaten. Daarna wordt de deelnemers gevraagd om de betekenis te koppelen aan het juiste element. Tot slot geeft de participant bij de (v) representatievoorkeur taken bij een gegeven betekenis met een slider aan naar welke PGA-notatie de voorkeur gaat. Hierbij wordt gebruik gemaakt van schaal van Cook et al. (2001), die loopt van 0 tot 100.

3.2.3. Operationele procedures

Tijdens het experiment zullen de reeds in het theoretisch kader beschreven procedurestappen worden gevolgd, die zijn in te delen in de volgende fasen: (i) inleiden experiment; (ii) uitvoeren experiment; (iii) afsluiten experiment en verzamelen van experimentele resultaten. Figuur 1 bevat een meer gedetailleerde beschrijving van procedurestap (ii) met de reeds beschreven experimentele taken.



Figuur 1: Volgordelijkheid experimentele taken tijdens het uitvoeren van het experiment

Omdat de standaardcommunicatie binnen het casusbedrijf digitaal is, worden zowel het experiment als de dataverzameling volledig digitaal gedaan. Het voordeel hiervan is dat de invloed van de onderzoeker op de onderzoeksresultaten wordt verkleind, wat ten goede komt aan de betrouwbaarheid van de onderzoeksresultaten. De tijdens het experiment gegeven informatie aan de participanten is immers altijd gelijk, alsook de kans dat het antwoord van participant wordt beïnvloed door de vraagwijze van de onderzoeker. Ook de vastlegging van de antwoorden is niet afhankelijk van de interpretatie van de onderzoeker.

3.2.4. Experimentele Groepen: Between- vs. Within-Subjects

In het onderzoek wordt vanwege de verschillende typen experimentele taken gebruik gemaakt van een combinatie van between-subjects en within-subjects design. Voor de notatieassociatie-, begrips- en herinneringstaken wordt een between-subjects design gebruikt, omdat de respondent per element ofwel de 'originele' ofwel de 'nieuwe' notatie te zien krijgt. Hierdoor zijn de onderzoeksresultaten van beide notaties met elkaar te vergelijken, waarbij het voor de hand ligt dat eventuele verschillen in intuïtiviteit verklaard kunnen worden door het verschil in notatieversie. Voor de representatievoorkeur taken is dit echter niet mogelijk. Participanten vergelijken immers beide notaties op een schaal van 0 tot 100 en worden hierbij blootgesteld aan zowel de 'originele' als de 'nieuwe' notatie. Daarom zal voor de representatievoorkeur taken gebruikt worden gemaakt van een within-subjects design. Een logisch gevolg van de between-subjects design is dat de groep participanten in twee subgroepen met gelijke grootte worden verdeeld met voor elke groep een eigen survey. Anders zijn de onderzoeksresultaten van beide groepen immers niet met elkaar te vergelijken. Om te voorkomen dat participanten inzicht krijgen in welke elementen nu precies tot de 'originele' of 'nieuwe' versie behoren - en dit van invloed kan zijn op hoe de representatievoorkeur taken worden

gemaakt - worden de originele en nieuwe elementen in de notatieassociatie-, begrips- en herinneringstaken gemixt aangeboden met in elke survey versie voor elke taak dezelfde combinatie van 'originele' en 'nieuwe' elementen. Met dit mogelijk ongewenste effect is ook rekening gehouden bij de representatievoorkeur taken, door tijdens de taak bij de schaal van 0 tot 100 van versie te wisselen: voor een deel van de taken staat aan de linkerkant van de schaal (i.e. 0) de 'originele' en aan de rechterkant (i.e. 100) de 'nieuwe' versie; en bij een deel van de taken staan de versies omgedraaid: aan de linkerkant (i.e. 0) de 'nieuwe' versie en aan de rechterkant (i.e. 100) de 'originele' versie.

Om ondanks de beperkte schaal (i.e. initieel 10 respondenten) de kracht van de experimentele test te maximaliseren, is het normaliter aan te raden om een within-subjects design te gebruiken (Saunders et al., 2019). Echter, gezien de gebruikte experimentele taken (i.e. notatieassociatie-, begrips- en herinneringstaken) is het noodzakelijk om deels ook een between-subjects design te hanteren. Om de kracht van het experiment te vergroten is er daarom voor gekozen om het aantal deelnemende respondenten te verhogen naar 15. Daarnaast zal hetzelfde experiment tegelijkertijd worden uitgevoerd in twee separate onderzoeken, bij twee andere casusorganisaties, waardoor het aantal deelnemende respondenten wordt verhoogd naar 45.

3.2.5. Hypothesen

Roelens and Bork (2020) stellen dat de in hun onderzoek uitgevoerde experimentele evaluatietechniek tot een meer intuïtieve PGA-notatie geleid zou moeten hebben. Daarentegen concluderen zij ook dat een vervolgevaluatie nodig is om vast te stellen of dit ook daadwerkelijk zo is. Hiertoe ligt het voor de hand om hypothesen te formuleren die per element 'i' (i.e. value proposition, internal goal, competence, customer goal, en de combinatie van importance en value stream) en per experimentele between-subjects taak 'j' (i.e. notatieassociatie-, begrips-, herinneringstaak) en within-subjects taak 'k' (i.e. representatievoorkeur taak) test of de 'nieuwe' notatie in vergelijking met de 'originele' notatie betere resultaten laat zien op de objectieve (i.e. interpretatieve effectiviteit en -efficiëntie) en subjectieve (i.e. representatievoorkeur) maatstaven met betrekking tot de intuïtieve begrijpelijkheid. Om per element een uitspraak te kunnen doen over in hoeverre de 'nieuwe' notatie als intuïtiever wordt beoordeeld dan de 'originele' ligt het eveneens voor de hand om hypothesen te formuleren die per element 'i' testen of het totale construct intuïtieve begrijpelijkheid (i.e. geaggregeerde score van interpretatieve effectiviteit, -efficiëntie en representatievoorkeur) voor de 'nieuwe' notatie hoger scoort in vergelijking met de 'originele'. Zodoende zijn de volgende vier soorten hypothesen te onderscheiden:

- $H_{\text{effectiviteit}_{ij}}$: de interpretatieve effectiviteit van de nieuwe PGA-notatie voor element i tijdens taak j is hoger dan die van de originele PGA-notatie voor element i tijdens taak j
- $H_{\text{efficiëntie}_{ij}}$: de interpretatieve efficiëntie van nieuwe PGA-notatie voor element i tijdens taak j is hoger dan die van de originele PGA-notatie voor element i tijdens taak j
- $H_{\text{representatievoorkeur}_{ik}}$: de representatievoorkeur van de nieuwe PGA-notatie voor element i tijdens taak k is hoger dan die van de originele PGA-notatie voor element i tijdens taak k
- $H_{\text{intuïtiviteit}_i}$: de intuïtieve begrijpelijkheid (i.e. geaggregeerde score van interpretatieve effectiviteit, -efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element i is hoger dan die van de originele PGA-notatie voor element i

Zie bijlage 7 voor een totaaloverzicht met alle (i.e. 42) hypothesen: 16 voor interpretatieve effectiviteit, 16 voor interpretatieve efficiëntie, 5 voor representatievoorkeur en 5 voor intuïtieve begrijpelijkheid.

3.2.6. Participanten

Het experiment wordt gedaan bij vijftien business users in een praktische bedrijfscontext. Er is specifiek voor dit type participant gekozen, omdat dit een positief effect zou moeten hebben op de betrouwbaarheid van het onderzoek. Door Roelens and Bork (2020) is immers aangegeven dat het gebruik van studenten als participanten in hun onderzoek wordt gezien als een inherente beperking van de betrouwbaarheid van de onderzoeksresultaten, omdat studenten geen daadwerkelijke eindgebruikers van de PGA-notatie zijn. Tevens zou de keuze voor deze participanten de externe validiteit ten goede moeten komen, omdat de resultaten hierdoor beter te veralgemenen zijn dan bij participanten die geen daadwerkelijke eindgebruikers zijn.

De praktische bedrijfscontext waarin het onderzoek zal plaatsvinden betreft een Nederlandse organisatie in de financiële sector, die zich vooral richt op het financieren van hypotheek, beheren van spaargeld en aanbieden van betaalrekeningen. De organisatie kent een EA-raamwerk waarin bedrijfsdoelen zijn beschreven en processen zijn gemodelleerd. Echter, deze zijn niet expliciet aan elkaar gekoppeld en op elkaar uitlijnd. Hierdoor is binnen de praktische bedrijfscontext dus wel kennis en ervaring aanwezig over doelen en processen, maar ontbreekt die over het koppelen en uitlijnen van beide. Hierdoor kunnen participanten in de casusorganisatie als daadwerkelijke eindgebruikers worden gezien en kan de praktische bedrijfscontext als geschikt worden bestempeld.

Aangezien de definitie van intuïtieve begrijpelijkheid gaat over het onmiddellijk begrijpen van de notatie, zonder voorkennis en/of training (Jošt et al., 2016), is het van belang om participanten te selecteren zonder diepgaand domeinbegrip (i.e. strategic fit) en/of modelleerervaring. Wel zou de participant enigszins domeinbegrip en modelleerervaring moeten hebben, om de termen en visualisaties te kunnen begrijpen. De PGA-notatie betreft tenslotte een DSML, die bedoeld is voor een specifiek toepassingsdomeinen die verrijkt zijn met termen en visualisaties waarmee eindgebruikers bekend zijn (Frank, 2010). Hiertoe worden participanten geselecteerd uit het middenmanagement, waarvan wordt verwacht dat deze aan de gestelde eisen voldoen. Enerzijds omdat dit type participant gezien kan worden als een realistische eindgebruiker, aangezien deze stakeholders toegang hebben tot de noodzakelijke informatie om het model op te kunnen bouwen, alsook het strategische belang van verschillende processen in kunnen schatten en belang hebben bij het verbeteren van strategic fit. Anderzijds vanuit de aanname dat het middenmanagement meer domeinbegrip en modelleerervaring heeft dan operationele medewerkers, maar minder dan het seniormanagement.

3.3. Gegevensanalyse

Bij het testen van hypothesen zou een kwantitatieve gegevensanalyse plaats moeten vinden op de experimentele onderzoeksresultaten, waarbij met statistische tests wordt beoordeeld in hoeverre de resultaten door toeval tot stand zijn gekomen (i.e. significance testing) (Saunders et al., 2019). Afhankelijk van een reeks dataeigenschappen kan worden bepaald welke statistische tests voor dit onderzoek relevant zijn. Allereerst is van belang welk type data (i.e. categorische of numerieke) de onafhankelijke variabele bevat. In dit onderzoek bevat de onafhankelijke variabele notatie categorische data en bestaat uit twee groepen, de 'originele' en 'nieuwe' notatie. Afhankelijk van het type experimenteel design (i.e. between-subjects of within-subjects) is vervolgens sprake van ongepaarde of gepaarde data. Met ongepaarde data wordt bedoeld dat een vergelijking wordt gemaakt tussen verschillende variabelen, bij gepaarde data tussen metingen van dezelfde variabelen (Studiedata, 2020). Bij de variabelen waarvoor een between-subjects design wordt gebruikt (i.e. interpretatieve effectiviteit, -efficiëntie en intuïtieve begrijpelijkheid), wordt een vergelijking gemaakt

tussen verschillende variabelen (i.e. 'originele' en 'nieuwe' notatie). Daarom is hier sprake van ongepaarde data. Bij de variabele waar een within-subjects design wordt gebruikt (i.e. representatievoorkeur) wordt een vergelijking gemaakt tussen metingen van dezelfde variabelen. Alle participanten worden immers blootgesteld aan zowel de 'originele' als de 'nieuwe' notatie. Hierdoor is deze data samen te nemen en is er sprake van gepaarde data. Vervolgens zijn - afhankelijk van het type data van de afhankelijke variabelen - twee groepen statistische tests te onderkennen: parametrische en non-parametrische tests. Parametrische tests worden gebruikt bij numerieke (i.e. interval en schaal) data en non-parametrische tests bij categorische (i.e. dichotome, nominale en ordinale) data. In dit onderzoek bevatten de afhankelijke variabelen numerieke en categorische data: de interpretatieve efficiëntie (i.e. tijd in seconden), intuïtieve begrijpelijkheid en representatievoorkeur bevatten numerieke (i.e. schaal) data; de interpretatieve effectiviteit (i.e. aantal correcte antwoorden) bevat categorische (i.e. ordinale) data. In geval van numerieke data moet bij het kiezen van de juiste hypothesetest ook rekening gehouden worden met de normaliteit van de data, waarbij wordt gekeken in hoeverre de data symmetrisch geclusterd is rondom het gemiddelde van de variabele. In dit onderzoek zal de normaliteit worden getest met behulp van de Shapiro-Wilk test. Bij ongepaarde variabelen met numerieke data (i.e. interpretatieve efficiëntie) zal in geval van normaal verdeelde data gebruik worden gemaakt van de parametrische Independent-samples T test. Bij numerieke, niet-normaal verdeelde (i.e. interpretatieve efficiëntie), en categorische (i.e. interpretatieve effectiviteit) data zal gebruik worden gemaakt van de non-parametrische Wilcoxon-Mann-Whitney Test. In geval van gepaarde numerieke data (i.e. representatievoorkeur) zal afhankelijk van de normaliteit ook worden gekozen voor parametrische of non-parametrische tests: bij normaal verdeelde data wordt de parametrische One-Sample T-test gebruikt, voor niet-normaal verdeelde data de non-parametrische One-sample Wilcoxon Signed Rank Test. Tabel 15 bevat een samenvatting van voor dit onderzoek relevante statistische tests.

Tabel 15: Voor dit onderzoek relevante statistische tests

Onafhankelijke variabele	Soort data	Gepaard/ ongepaard	Afhankelijke variabele	Soort data	Normaliteit	Statistische toets
Notatie	Categorisch, 2 groepen	Ongepaard	Interpretatieve efficiëntie	Numeriek, Schaal	Normaal	Independent-samples T test
			Intuïtieve begrijpelijkheid		Niet normaal	Wilcoxon-Mann-Whitney Test
			Interpretatieve effectiviteit	Categorisch, Ordinaal	Nvt	
		Gepaard	Representatievoorkeur	Numeriek, Schaal	Normaal	One-Sample T-test
					Niet normaal	One-sample Wilcoxon Signed Rank Test

3.4. Reflectie t.a.v. validiteit, betrouwbaarheid en ethische aspecten

3.4.1. Constructvaliditeit

Constructvaliditeit zegt iets over tot welke hoogte het meetinstrument meet wat het moet meten (Saunders et al., 2019). Dit wordt geborgd door gebruik te maken van de in het theoretisch kader beschreven, op voorgaand onderzoek gebaseerde afhankelijke (i.e. notatie) en onafhankelijke (i.e. interpretatieve efficiëntie, interpretatieve effectiviteit, representatievoorkeur) variabelen waarbij ook gebruik gemaakt wordt van de hierbij gebruikte, bestaande, reeds gevalideerde evaluatie-/meetmethodes (i.e. notatieassociatie-; begrips-; herinnerings-; en representatievoorkeur taken).

3.4.2. Interne validiteit

Interne validiteit heeft te maken met de mate waarin een verondersteld causaal verband daadwerkelijk optreedt. Hiervoor zou moeten worden gepoogd om verstorende (i.e. derde) factoren zoveel mogelijk uit te sluiten (Saunders et al., 2019). Zo zal het experiment volledig digitaal worden aangeboden, zonder menselijke interactie. Dit heeft als voordeel dat de tijdens het experiment gegeven informatie aan de participanten altijd gelijk is. Daarnaast kan het antwoord van de participant niet worden beïnvloed door de vraagwijze van de onderzoeker en is ook de vastlegging van de antwoorden niet afhankelijk van de interpretatie van de afnemer. Een ander voordeel is dat het digitaal aanbieden nauw doet aansluiten bij de standaardcommunicatie binnen de praktische bedrijfscontext waarbinnen het onderzoek plaatsvindt, zodat het medium waarmee het experiment wordt aangeboden niet als verstorende factor kan worden gezien. Eveneens wordt de interne validiteit bevorderd, door te bewerkstelligen dat het experiment door alle participanten binnen een beperkte tijdsperiode wordt gemaakt, waardoor in grote mate sprake zal zijn van dezelfde omstandigheden waaronder het experiment is uitgevoerd. Ook is de deelname aan het onderzoek vrijwillig en zal er geen compensatie worden verstrekt.

Aangezien het experimenteel design in dit onderzoek deels een between-subjects design bevat, wordt in het onderzoek gebruik gemaakt van twee survey versies, zodat de onderzoeksresultaten over de intuïtiviteit van de 'originele' en 'nieuwe' notatie met elkaar te vergelijken zijn. Om te voorkomen dat participanten bij het uitvoeren van de representatietaken worden beïnvloed door het hebben van inzicht in welke elementversie (i.e. 'origineel' of 'nieuw') zij in de voorgaande taken hebben gezien, worden de elementen in de notatieassociatie-, begrips- en herinneringstaken gemixt aangeboden met in elke survey versie, voor elke taak dezelfde combinatie van 'originele' en 'nieuwe' elementen. Bij de representatievoorkeur taken zelf is hiermee rekening gehouden, door de richting van de schaal te wisselen: voor een deel van de taken staat aan de linkerkant van de schaal (i.e. 0) de 'originele' en aan de rechterkant (i.e. 100) de 'nieuwe' versie en bij een deel van de taken andersom. Om uit te sluiten dat een gebrek aan specifieke werksector kennis een invloedfactor is voor het fout beantwoorden van de begripstaken, is gekozen voor een model van een organisatie waarvan wordt verwacht dat participanten hiermee bekend zijn: een online marktplaats voor het aanbieden van eten door restaurantpartners en het bestellen ervan door klanten.

3.4.3. Externe validiteit

Externe validiteit gaat over in hoeverre onderzoeksresultaten gegeneraliseerd kunnen worden (Saunders et al., 2019). De verwachting is dat de participanten in dit onderzoek (i.e. business users in een praktische bedrijfscontext) meer overeenkomen met de daadwerkelijke eindgebruikers van de PGA-notatie dan de participanten (i.e. masterstudenten) uit voorgaand onderzoek, wat de externe validiteit van de onderzoeksresultaten ten goede zou moeten komen. Specifiek zullen de resultaten van dit onderzoek te veralgemenen zijn voor het middenmanagement van Nederlandse organisaties in de financiële sector.

3.4.4. Betrouwbaarheid

Bij betrouwbaarheid wordt gekeken naar tot welke hoogte resultaten reproduceerbaar zijn (Saunders et al., 2019). De betrouwbaarheid van het onderzoek wordt geborgd door het leveren van een gedetailleerde beschrijving van de methodische opzet (i.e. hoofdstuk 3) en de tijdens het experiment gebruikte survey (i.e. bijlage 5), alsook van de resultaten (i.e. hoofdstuk 4), waardoor het aannemelijk is dat een replicatie van het experiment tot dezelfde resultaten zullen leiden. Tevens zal de betrouwbaarheid van de onderzoeksresultaten in vergelijking met voorgaand onderzoek worden vergroot door voor participanten te kiezen die overeenkomen met daadwerkelijke eindgebruikers (i.e. business users in een praktische bedrijfscontext). Dit werd door (Roelens & Bork, 2020) immers als inherente beperking aangemerkt voor de betrouwbaarheid van hun onderzoeksresultaten. Ook zal het experiment digitaal worden afgenomen, waardoor de onderzoeker geen ongewenste invloed uit kan oefenen op de onderzoeksresultaten en menselijke fouten bij vastlegging worden voorkomen. Een ander voordeel hiervan is dat digitale communicatie bij de casusorganisatie ook de standaard is, zodat een meer realistisch gebruik wordt nagebootst dan met een papieren afname. Ook wordt de betrouwbaarheid vergroot door bij de taken precies de informatie te verstrekken die nodig is om een min of meer realistisch gebruik te simuleren. Tot slot zal de betrouwbaarheid van de resultaten worden vergroot door gebruik te maken van een combinatie van objectieve en subjectieve variabelen. Hierdoor treedt triangulatie op, waardoor de resultaten betrouwbaarder zijn, dan wanneer gebruik gemaakt wordt van slechts één bron.

3.4.5. Ethische Aspecten

Voor wat betreft de ethische aspecten wordt in ogeschouw genomen in hoeverre het gedrag van de onderzoeker verantwoord is in relatie tot diegenen die onderwerp worden van het onderzoek, of die daardoor worden beïnvloed (Saunders et al., 2019). Hiermee wordt rekening gehouden door volledig vrijwillige deelname (i.e. informed consent), waarbij is voorzien in de mogelijkheid dat de participant tijdens het experiment op elk willekeurig moment kan stoppen. De verzamelde data zal vertrouwelijk worden behandeld en de participantnamen alsook de bedrijfsnaam van de casusorganisatie zullen worden geanonimiseerd.

4. Resultaten

Dit hoofdstuk bevat een beschrijving van de uitvoering van het onderzoek (i.e. participanten en experimentele opzet) en de analyse van de onderzoeksresultaten (i.e. vergelijkende evaluatieresultaten binnen de praktische bedrijfscontext) waarmee een antwoord wordt gegeven op de deelonderzoeksvraag: *'In hoeverre wordt de 'nieuwe' PGA-notatie als intuïtiever beoordeeld in vergelijking met de 'originele' notatie?'*.

4.1. Participanten en experimentele opzet

Dit onderdeel voorziet in een korte beschrijving van de uitvoering van het onderzoek, alsook de mate waarin dit heeft plaatsgevonden volgens het oorspronkelijke plan zoals beschreven in het voorgaande hoofdstuk.

4.1.1. Participanten

In totaal zijn 37 personen (i.e. middenmanagers) in de praktische bedrijfscontext (i.e. Nederlandse organisatie in de financiële sector) benaderd, waarvan er 15 daadwerkelijk aan het onderzoek hebben meegedaan (i.e. response rate van 43%). Zoals reeds beschreven in hoofdstuk 3 is het experiment tegelijkertijd uitgevoerd in twee separate onderzoeken, bij twee andere casusorganisaties, die zorgen voor 29 extra participanten. Van de totale groep deelnemers (i.e. $15 + 29 = 44$ participanten) is 68% werkzaam in de private sector en 27% in de publieke sector. Van de participanten is de gemiddelde leeftijd 45 jaar, 75% is man, 23% vrouw en 2% heeft geantwoord met 'X'. Als hoogst genoten opleiding geeft 61% aan een master-graad te bezitten, 34% een bachelor. Hoewel participanten geen ervaring hebben met de PGA-notatie geeft 66% aan bekend te zijn met conceptuele modelleertalen voor Business Process Management, 50% voor Requirements Engineering en 43% voor Enterprise Architecture - 27% geeft aan geen enkele modelleertaal te kennen. Participanten hebben in de afgelopen 5 jaar gemiddeld tussen de 11 en 20 conceptuele modellen gelezen, tussen de 1 en 10 gemaakt en bestempelen hun modelleerervaring als 'gemiddeld'. Echter bij de modelleerexpertise en -ervaring zijn enkele opvallende verschillen te zien tussen eigen dataset en die van de separate onderzoeken. Zo geven participanten van de eigen dataset vaker aan met geen enkele modelleertaal ervaring te hebben: 40% (i.e. 6 participanten) van de eigen dataset ten opzichte van 21% (i.e. 6 participanten) in de andere datasets. Daarnaast heeft in de eigen dataset slechts 47% (i.e. 7 participanten) ervaring met BPM-modelleertalen, ten opzichte van 75% (i.e. 22 participanten) in de andere datasets. Ook geeft in de eigen dataset 27% (i.e. 4 participanten) aan geen enkele modelleerexpertise te hebben ten opzichte van 14% (i.e. 5 participanten) van andere datasets. Deze verschillen kunnen er mogelijk toe leiden dat er verschillen te zien zijn in de onderzoeksresultaten en dat deze hierdoor wellicht als minder betrouwbaar kunnen worden bestempeld.

4.1.2. Experimentele opzet

In het onderzoek is gebruik gemaakt van een online survey (i.e. Limesurvey) die via e-mail met de potentiële participanten is gedeeld, waarin is aangegeven dat deze gedurende drie weken ingevuld kon worden. Voorafgaand hieraan zijn pretests uitgevoerd met personen die later niet hebben deelgenomen aan het daadwerkelijke onderzoek. De survey bevat demografische vragen en de volgende experimentele taken: notatieassociatie taken, begripstaken, herinneringstaken en representatievoorkeur taken. Zoals in hoofdstuk 3 staat beschreven zijn de participanten verdeeld in

twee subgroepen met elk een eigen survey (i.e. versie A of B) waarin de notaties zijn gewisseld met voor alle taken van een bepaalde groep een specifieke combinatie van 'originele' en 'nieuwe' elementen, zie tabel 16. De versies zijn automatisch door de online tool aan de participanten toegewezen op basis van het geboortjaar (i.e. even geboortjaar kreeg versie A, oneven geboortjaar kreeg versie B). Dit heeft geleid tot een evenwichtige verdeling: van de respondenten heeft 55% versie A ingevuld, 45% versie B. Gemiddeld duurde het invullen van de survey 18 minuten.

Tabel 16: verdeling oude en nieuwe notatie over survey versie A en B

Element	Notatie Versie A / VersionCode 0	Notatie Versie B / VersionCode 1
Value proposition	Old	New
Importance	Old	New
Internal goal	New	Old
Competence	New	Old
Customer goal	Old	New
Value stream	Old	New

Na het verstrijken van de deadline zijn de surveyresultaten digitaal opgeslagen en wanneer van toepassing handmatig in excel gecodeerd, zodat statistische toetsen konden worden uitgevoerd in SPSS. Deze zijn uitgevoerd per element op taakniveau voor zowel de eigen als de volledige dataset. Nadien zijn op beide datasets ook per element analyses uitgevoerd op het hele construct intuïtiviteit (i.e. geaggregeerde score van interpretatieve effectiviteit, -efficiëntie en representatievoorkeur).

4.2. Analyse resultaten per element op taakniveau

4.2.1. Resultaten eigen dataset

In de eigen dataset (i.e. 15 respondenten) lijkt voor de elementen op taakniveau in een drietal gevallen de 'nieuwe' notatie als intuïtiever te worden beoordeeld in vergelijking met de 'originele', zie tabel 17 en voor de uitgebreide analyse bijlage 8. Zo scoort de representatievoorkeur voor de 'nieuwe' notaties voor de elementen value proposition en competence tijdens de representatievoorkeur taak significant beter dan de 'originele' op een significantieniveau van 5 procent. Dat wil zeggen dat de kans meer dan 95% is dat het scoreverschil tussen beide versies niet op toeval berust en dus dat de 'nieuwe' notatie waarschijnlijk als beter wordt beoordeeld op representatievoorkeur in vergelijking met de 'originele'. Ditzelfde geldt voor de interpretatieve efficiëntie van de 'nieuwe' notatie voor element internal goal tijdens de herinneringstaak. Dit betekent dus dat 2 van de 5 'nieuwe' notaties bij representatievoorkeur beter scoren dan de 'originele', 1 van de 16 bij interpretatieve efficiëntie en 0 van de 16 bij de interpretatieve effectiviteit.

4.2.2. Resultaten volledige dataset

Op het niveau van de volledige dataset (i.e. 44 respondenten) is bij een zestal elementen de 'nieuwe' notatie als intuïtiever beoordeeld dan de 'originele'. Hierin zijn twee overeenkomstige uitkomsten te zien in vergelijking met de eigen dataset (i.e. 15 respondenten): de 'nieuwe' notaties scoren voor de representatievoorkeur van element value proposition en de interpretatieve efficiëntie van de internal goal bij de herinneringstaak eveneens significant beter dan de 'originele', zie tabel 17 en voor de uitgebreide analyse bijlage 9. Tevens zijn er afwijkingen te zien. Zo scoort in de totale dataset ook de interpretatieve effectiviteit van de 'nieuwe' notatie van value proposition tijdens de notatieassociatie taak hoger dan de 'originele'. In tegenstelling tot de eigen dataset scoort de representatievoorkeur van de 'nieuwe' notatie van element competence niet hoger dan de 'originele', maar is dat wel het geval voor de interpretatieve efficiëntie tijdens de herinneringstaak. Ook scoort de interpretatieve efficiëntie van de 'nieuwe' notatie van element importance tijdens de begripstaak beter dan het 'origineel' alsook de representatievoorkeur voor de 'nieuwe' notatie van element customer goal. Dit betekent dus dat 2 van de 5 'nieuwe' notaties bij representatievoorkeur beter scoren dan de 'originele', 3 van de 16 bij interpretatieve efficiëntie en 1 van de 16 bij de interpretatieve effectiviteit.

Tabel 17: resultaten per element op taakniveau van de eigen en totale dataset

Nulhypothese	Statistische test <i>boven: eigen set; onder: totale set</i>	Asymp. Sig. (1-tailed) <i>boven: eigen set; onder: totale set</i>	Significant <i>boven: eigen set; onder: totale set</i>
H <u>representatievoorkeur</u> : de <u>representatievoorkeur</u> van de nieuwe PGA-notatie voor element value proposition tijdens de representatievoorkeur taak is niet hoger dan die van de originele PGA-notatie voor element value proposition tijdens de representatievoorkeur taak	Wilcoxon Test, Sign Test	0.02	ja
	Wilcoxon Test, Sign Test	0.0005	ja
H <u>representatievoorkeur</u> : de <u>representatievoorkeur</u> van de nieuwe PGA-notatie voor element competence tijdens de representatievoorkeur taak is niet hoger dan die van de originele PGA-notatie voor element competence tijdens de representatievoorkeur taak	One-Sample T-test	0.02	ja
	Wilcoxon Test, Sign Test	0.269	nee
H <u>representatievoorkeur</u> : de <u>representatievoorkeur</u> van de nieuwe PGA-notatie voor element customer goal tijdens de representatievoorkeur taak is niet hoger dan die van de originele PGA-notatie voor element customer goal tijdens de representatievoorkeur taak	One-Sample T-test	0.48	nee
	Wilcoxon Test, Sign Test	0.021	ja
H <u>efficiëntie</u> : de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element internal goal tijdens de herinneringstaak is niet hoger dan die van de originele PGA-notatie voor element internal goal tijdens de herinneringstaak	Wilcoxon-Mann-Whitney Test	0.01	ja
	Wilcoxon-Mann-Whitney Test	0.045	ja
H <u>efficiëntie</u> : de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element competence tijdens de herinneringstaak is niet hoger dan die van de originele PGA-notatie voor element competence tijdens de herinneringstaak	Independent-samples T test	0.05	nee
	Wilcoxon-Mann-Whitney Test	0.0495	ja
H <u>efficiëntie</u> : de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element importance tijdens de begripstaak is niet hoger dan die van de originele PGA-notatie voor element importance tijdens de begripstaak	Independent-samples T test	0.23	nee
	Wilcoxon-Mann-Whitney Test	0.004	ja
H <u>effectiviteit</u> : de <u>interpretatieve effectiviteit</u> van de nieuwe PGA-notatie voor element value proposition tijdens de notatieassociatie taak is niet hoger dan die van de originele PGA-notatie voor element value proposition tijdens de notatieassociatie taak	Wilcoxon-Mann-Whitney Test	0.18	nee
	Wilcoxon-Mann-Whitney Test	0.026	ja

4.3. Analyse geaggregeerde resultaten intuïtiviteit per element

In tabel 18 is te zien dat op het niveau van zowel de eigen als de volledige dataset voor wat betreft het hele construct intuïtiviteit (i.e. geaggregeerde score van interpretatieve effectiviteit, -efficiëntie en representatievoorkeur) voor geen enkel element de 'nieuwe' notatie als intuïtiever wordt beoordeeld in vergelijking met de 'originele'. Dit betekent dus dat 0 van de 5 'nieuwe' notaties op construct intuïtiviteit beter scoren dan de 'originele', zie bijlage 10 en 11 voor de uitgebreide analyse van respectievelijk de eigen en de totale dataset.

Tabel 17: Resultaten construct intuïtieve begripelijkheid eigen en totale dataset

Nulhypothese	Statistische test <i>boven: eigen set; onder: totale set</i>	Asymp. Sig. (1-tailed) <i>boven: eigen set; onder: totale set</i>	Significant <i>boven: eigen set; onder: totale set</i>
H_intuïtiviteit: de <u>intuïtieve begripelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, -efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element internal goal is niet hoger dan die van de originele PGA-notatie voor element internal goal	Wilcoxon-Mann-Whitney Test	0.41	nee
	Wilcoxon-Mann-Whitney Test	0.16	nee
H_intuïtiviteit: de <u>intuïtieve begripelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, -efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element value proposition is niet hoger dan die van de originele PGA-notatie voor element value proposition	Independent-samples T test	0.1525	nee
	Independent-samples T test	0.3165	nee
H_intuïtiviteit: de <u>intuïtieve begripelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, -efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element competence is niet hoger dan die van de originele PGA-notatie voor element competence	Independent-samples T test	0.407	nee
	Independent-samples T test	0.197	nee
H_intuïtiviteit: de <u>intuïtieve begripelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, -efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element value stream is niet hoger dan die van de originele PGA-notatie voor element value stream	Independent-samples T test	0.4755	nee
	Wilcoxon-Mann-Whitney Test	0.40	nee
H_intuïtiviteit: de <u>intuïtieve begripelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, -efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element customer goal is niet hoger dan die van de originele PGA-notatie voor element customer goal	Wilcoxon-Mann-Whitney Test	0.35	nee
	Wilcoxon-Mann-Whitney Test	0.13	nee

4.4. Kwalitatieve analyse

4.4.1. Reacties tijdens en na experiment

Tijdens en na het uitvoeren van het experiment is in het eigen onderzoek door 12 (i.e. 32%) van de in totaal 37 benaderde, potentiële participanten aangegeven dat de survey als moeilijk werd ervaren. Dit lijkt ook te worden bevestigd door het aantal (i.e. 19) deels ingevulde surveys. Over de survey werd aangemerkt dat niet altijd duidelijk was wat tijdens de taken van de participant werd verwacht. Daarnaast werd aangegeven dat de modelleertaal niet goed werd begrepen, omdat deze veel jargon bevatte waarmee respondenten niet bekend waren. Opvallend is dat deze reacties niet aanwezig waren bij de twee separate onderzoeken, bij andere casusorganisaties. Dit verschil lijkt ook zichtbaar te zijn in het aantal statistische tests waarin de 'nieuwe' notatie als intuïtiever wordt beoordeeld in vergelijking met de 'originele': 3 in de eigen dataset ten opzichte van 6 in de totale dataset. Een verklaring hiervoor zou gevonden kunnen worden in de reeds beschreven verschillen (i.e. modelleerexpertise en -ervaring) tussen de participanten in de eigen dataset en die van de separate onderzoeken. Zo lijken participanten van de eigen dataset aanzienlijk vaker geen enkele modelleertaal ervaring te hebben; minder ervaring te hebben met BPM-modelleertalen; en vaker geen enkele modelleerexpertise te hebben ten opzichte van participanten in de andere datasets. Dit kan negatieve consequenties hebben voor de betrouwbaarheid van de onderzoeksresultaten alsook de interne validiteit.

4.4.2. Kwalitatieve feedback representatievoorkeur

In het experiment is bij de representatievoorkeur taak de optie geboden om een tekstuele (i.e. kwalitatieve) toelichting te geven op de representatievoorkeur. Om de betrouwbaarheid van de onderzoeksresultaten te vergroten kan deze kwalitatieve data worden gebruikt om doormiddel van triangulatie te beoordelen in hoeverre deze de resultaten van de statistische testen over de representatievoorkeur hypothesen ondersteunen. Het uitvoeren van een thematische analyse voor het identificeren van thema's en patronen in de kwalitatieve data is hiervoor een geschikte methode (Saunders et al., 2019). Hierbij worden data met een gelijksoortige betekenis gecodeerd (i.e. een woord of korte zin) zodat thema's onderling met elkaar kunnen worden vergeleken. Aangezien gebruik is gemaakt van een deductieve onderzoekaankpak worden per element twee a priori codes (i.e. één voor de nieuwe en één voor de originele) gebruikt die zijn afgeleid van de gebruikte symbolen in de oude en nieuwe notaties. Hierbij wordt specifiek aandacht besteedt aan de mate waarin er sprake is van contrasterende meningen over de aangebrachte verbeteringen. Om de kwalitatieve data als ondersteunend te zien voor de representatievoorkeur voor de 'nieuwe' notatie van een element, worden de volgende criteria toegepast: (i) minimaal 11 keer (i.e. 25% van de totale populatie) moet een specifiek thema worden genoemd in de kwalitatieve data; (ii) vervolgens moet dit thema minimaal twee-derde keer (i.e. 67%) gebruikt worden ten gunste van de nieuwe notatie. Per element moet minimaal één van de twee thema's aan het criterium voldoen om als ondersteunend te worden gezien voor de representatievoorkeur voor de 'nieuwe' notatie.

Tabel 18: Resultaten analyse kwalitatieve feedback representatievoorkeur

Element	Thema	% thema genoemd t.o.v. totale populatie (aantal keer genoemd)	% contrasterende meningen t.v.v. nieuwe notatie (aantal keer keuze voor nieuw;oud)	Ondersteunend t.v.v. representatie-voorkeur nieuwe notatie
Value proposition	Cadeau	41% (18)	83% (15;3)	ja
	Schaal	32% (14)	100% (14;0)	ja
Customer goal	Stakeholders	45% (20)	85% (17;3)	ja
	Handen	18% (8)	88% (7;1)	nee
Competence	Lamp	39% (17)	47% (8;9)	nee
	Podium	36% (16)	63% (10;6)	nee
Internal goal	Fabriek	45% (20)	60% (12;8)	nee
	Tandwielen	11% (5)	80% (4;1)	nee
Value stream/ Importance	Kleuren	48% (21)	52% (10;11)	nee
	Uitroeptekens	18% (8)	75% (2;6)	nee

In de resultaten van de analyse van de kwalitatieve data in tabel 19 is te zien dat de data alleen de representatievoorkeur voor de 'nieuwe' notatie voor de elementen value proposition en customer goal ondersteunen. De tabel kan als volgt worden gelezen: kolom 'Thema' bevat de a priori codes die zijn afgeleid van de gebruikte symbolen in de oude en nieuwe notaties; '% thema genoemd t.o.v. totale populatie' bevat het percentage dat weergeeft of wordt voldaan aan criterium (i), inclusief het aantal keer dat het thema werd genoemd; '% contrasterende meningen t.v.v. nieuwe notatie' geeft het percentage weer dat iets zegt over de mate waarin wordt voldaan aan criterium (ii), inclusief informatie over de verhouding tussen het aantal keren dat het thema is gebruikt als argument om de nieuwe notatie te kiezen, versus het aantal keren dat vanwege het thema juist voor de originele notatie is gekozen; 'Ondersteunend t.v.v. representatie-voorkeur nieuwe notatie' vertelt vervolgens per element, per thema wat de conclusie is, op basis van de gehanteerde criteria (i) en (ii). Zie voor de uitgebreide analyse bijlage 12.

5. Discussie, conclusies en aanbevelingen

In het onderzoek is op basis van methodologische literatuur een vervolgexperiment opgezet om met behulp van een vergelijkende evaluatie de intuïtieve begrijpelijkheid te evalueren tussen twee versies van de PGA-notatie (i.e. 'nieuwe' versus 'originele' versie) in een praktische bedrijfscontext. Tevens is op basis van de literatuur een reeks variabelen geoperationaliseerd en omgezet in hypothesen over de interpretatieve effectiviteit, interpretatieve efficiëntie, representatievoorkeur en het totale construct intuïtieve begrijpelijkheid. In dit deel van het onderzoek worden de onderzoeksresultaten uit hoofdstuk 4 samengevat en wordt vervolgens per element geanalyseerd in hoeverre de evaluatieresultaten de hypothesen uit hoofdstuk 3 ondersteunen. Vervolgens wordt gereflecteerd op de validiteit, betrouwbaarheid en ethische aspecten van het onderzoek. Daarna volgt een samenvatting van de belangrijkste conclusies, waarmee uiteindelijk een antwoord gegeven op de vraag over de mate waarin de 'nieuwe' PGA-notatie intuïtiever is dan de 'originele'. Daarnaast zal ook een uitspraak worden gedaan over de mate waarin de evaluatietechniek van Bork et al. (2019) heeft geleid tot significante verbeteringen in de intuïtiviteit van de PGA-notatie voor eindgebruikers in een praktische bedrijfscontext. Eveneens zal een besluit worden genomen over eventueel benodigde aanpassingen aan de experimentele opzet van dit onderzoek, waarmee tevens ook een antwoord wordt gegeven op de centrale onderzoeksvraag *'Hoe kan de intuïtiviteit van verschillende versies van een DSML-notatie empirisch met elkaar worden vergeleken in een praktische bedrijfscontext?'.* Tot slot volgen aanbevelingen voor de praktijk en voor vervolgonderzoek.

5.1. Discussie onderzoeksresultaten

5.1.1. Samenvatting kwantitatieve en kwalitatieve resultaten per element

Tabel 20 vat de kwantitatieve en kwalitatieve resultaten uit hoofdstuk 4 samen. Voor de kwantitatieve resultaten is per element aangegeven welke nulhypothese(n) volgens de statistische tests niet kunnen worden verworpen (i.e. -) en welke alternatieve hypothesen kunnen worden aanvaard (i.e. x) op basis van de eigen (i.e. E) en/of totale (i.e. T) dataset. De hypothesen zijn immers geformuleerd per taak (i.e. notatieassociatie-, begrips-, herinnerings-, en representatievoorkeur taak) en per objectieve variabele interpretatieve effectiviteit (i.e. effect); - efficiëntie (i.e. efficiënt); en subjectieve variabele representatievoorkeur (i.e. repv). Ook zijn hypothesen geformuleerd voor het totale construct intuïtieve begrijpelijkheid (i.e. geaggregeerd). Voor de kwalitatieve data is per element weergegeven of de kwalitatieve data (i.e. qual) uit de representatievoorkeur taak de alternatieve representatievoorkeur hypothese wel (i.e. Ja) of niet (i.e. Nee) ondersteunen.

Tabel 20: totaaloverzicht samenvatting kwantitatieve en kwalitatieve resultaten per element

Element	Notatieassociatie taak		Begripstaak		Herinneringstaak		Representatievoorkeur taak		Geaggregeerd
	effectief	efficiënt	effectief	efficiënt	effectief	efficiënt	repv	qual	
Value proposition	x (T)	-	-	-	-	-	x (T+E)	Ja	-
Customer goal	-	-	-	-	-	-	x (T)	Ja	-
Competence	-	-	-	-	-	x (T)	x (E)	Nee	-
Internal goal	-	-	-	-	-	x (T+E)	-	Nee	-
Value stream/Importance	-	-	-	x (T)	-	-	-	Nee	-

In de volgende hoofdstukken wordt per element (i.e. value proposition, customer goal, competence, internal goal en value stream/importance) op basis van de kwantitatieve en kwalitatieve onderzoeksresultaten beoordeeld in hoeverre de 'nieuwe' notatie als intuïtiever is beoordeeld in vergelijking met de 'originele'.

5.1.2. Value proposition

In de resultaten is te zien dat bij 2 van de 8 hypothesen de 'nieuwe' notatie beter wordt beoordeeld in vergelijking met de 'originele' notatie. De objectieve maatstaf interpretatieve effectiviteit bij de notatieassociatie taak scoort in de totale dataset beter; en bij de subjectieve maatstaf representatievoorkeur is dat zelfs het geval in zowel de eigen als de totale dataset. Dit laatste wordt ook ondersteund door de kwalitatieve data. Participanten geven immers aan voor de 'nieuwe' notatie te hebben gekozen, omdat het cadeau symbool in deze notatie met waarde werd geassocieerd alsook omdat bij de 'originele' notatie geen link werd gezien tussen de weegschaal en de betekenis van het element. Concluderend kunnen dus de twee nulhypothesen worden verworpen en de volgende alternatieve hypothesen worden aanvaard:

- H representatievoorkeur (totale + eigen dataset): de representatievoorkeur van de nieuwe PGA-notatie voor element **value proposition** tijdens de **representatievoorkeur taak** is hoger dan die van de originele PGA-notatie voor element value proposition tijdens de representatievoorkeur taak
- H effectiviteit (totale dataset): de interpretatieve effectiviteit van de nieuwe PGA-notatie voor element **value proposition** tijdens de **notatieassociatie taak** is hoger dan die van de originele PGA-notatie voor element value proposition tijdens de notatieassociatie taak

5.1.3. Customer goal

In de resultaten is te zien dat de 'nieuwe' notatie voor element customer goal bij 1 van de 8 hypothesen als beter wordt beoordeeld in vergelijking met de 'originele' notatie. De subjectieve maatstaf representatievoorkeur taak scoort in de totale dataset beter. Dit laatste wordt ook ondersteund door de kwalitatieve data: participanten geven immers aan voor de 'nieuwe' notatie te hebben gekozen, omdat in deze versie twee verschillende (i.e. business en non-business) stakeholders te zien zijn waarin de klantrelatie wordt herkend. Concluderend kan dus de nulhypothese worden verworpen en de volgende alternatieve hypothese worden aanvaard:

- H representatievoorkeur (totale dataset): de representatievoorkeur van de nieuwe PGA-notatie voor element **customer goal** tijdens de **representatievoorkeur taak** is hoger dan die van de originele PGA-notatie voor element customer goal tijdens de representatievoorkeur taak

5.1.4. Competence

Bij element competence is te zien dat bij 2 van de 8 hypothesen de 'nieuwe' notatie als beter wordt beoordeeld in vergelijking met de 'originele' notatie. De objectieve maatstaf interpretatieve efficiëntie bij de herinneringstaak scoort in totale dataset beter; en bij de subjectieve maatstaf representatievoorkeur in de eigen dataset is dit eveneens het geval. Overigens wordt dit laatste door de kwalitatieve data niet ondersteund. De contrasterende meningen van de participanten wijken immers niet genoeg van elkaar af. Concluderend kunnen dus de twee nulhypothesen worden verworpen en de volgende alternatieve hypothesen worden aanvaard:

- H representatievoorkeur (eigen dataset): de representatievoorkeur van de nieuwe PGA-notatie voor element **competence** tijdens de **representatievoorkeur taak** is hoger dan die van de originele PGA-notatie voor element competence tijdens de representatievoorkeur taak

- H efficiëntie (totale dataset): de interpretatieve efficiëntie van nieuwe PGA-notatie voor element **competence** tijdens de **herinneringstaak** is hoger dan die van de originele PGA-notatie voor element competence tijdens de herinneringstaak

5.1.5. Internal goal

In het geval van element internal goal is te zien dat bij 1 van de 8 hypothesen de 'nieuwe' notatie als beter wordt beoordeeld in vergelijking met de 'originele'. De objectieve maatstaf interpretatieve efficiëntie bij de herinneringstaak scoort in zowel de totale als de eigen dataset beter. Concluderend kan dus de nulhypothese worden verworpen en de volgende alternatieve hypothese worden aanvaard:

- H efficiëntie (totale en eigen dataset): de interpretatieve efficiëntie van nieuwe PGA-notatie voor element **internal goal** tijdens de **herinneringstaak** is hoger dan die van de originele PGA-notatie voor element internal goal tijdens de herinneringstaak

5.1.6. Value stream/Importance

De 'nieuwe' notatie voor element value stream/importance wordt bij 1 van de 8 hypothesen als beter wordt beoordeeld in vergelijking met de 'originele'. De objectieve maatstaf interpretatieve efficiëntie bij de begripstaak scoort in de totale dataset beter. Concluderend kan dus de nulhypothese worden verworpen en de volgende alternatieve hypothese worden aanvaard:

- H efficiëntie (totale dataset): de interpretatieve efficiëntie van nieuwe PGA-notatie voor element **importance** tijdens de **begripstaak** is hoger dan die van de originele PGA-notatie voor element importance tijdens de begripstaak

5.2. Reflectie validiteit en betrouwbaarheid

In dit deel van het onderzoek wordt gereflecteerd op de in hoofdstuk 3 geformuleerde voorgenomen beheersmaatregelen om de validiteit (i.e. construct-, intern- en extern-) en betrouwbaarheid van het onderzoek te borgen.

5.2.1. Constructvaliditeit

Zoals reeds in hoofdstuk 3 is beschreven zegt constructvaliditeit iets over tot welke hoogte het meetinstrument meet wat het moet meten (Saunders et al., 2019). Dit is in het onderzoek geborgd door gebruik te maken van de in het theoretisch kader beschreven, op voorgaand onderzoek gebaseerde afhankelijke en onafhankelijke variabelen. Ook is waar mogelijk gebruik gemaakt van bestaande, reeds gevalideerde evaluatie-/meetmethodes. Echter, voor een vergelijkende evaluatie van de intuïtieve begrijpelijkheid van twee versies van een notatie is geen reeds gevalideerd meetinstrumenten voorhanden, waardoor het noodzakelijk was om zelf een experiment te ontwerpen. Hoewel dit is gedaan op basis van theorie is dus niet met zekerheid te zeggen of het meetinstrument meet wat het moet meten.

Dit lijkt ook te worden bevestigd door de in tabel 20 samengevatte onderzoeksresultaten, waarin te zien is dat bij de kwantitatieve resultaten van de totale dataset slechts 6 (i.e. 15%) van de 40 alternatieve hypothesen kunnen worden aanvaard en 34 nulhypothesen kunnen worden verworpen. Van die 6 zijn er 2 (i.e. value proposition en customer goal) op de variabele representatievoorkeur; 3 op interpretatieve efficiëntie (i.e. competence en internal goal bij herinneringstaak; value stream/importance bij begripstaak); 1 op interpretatieve effectiviteit (i.e. value proposition bij notatieassociatie taak); en 0 op het totale construct intuïtieve begrijpelijkheid. Bij de eigen dataset is dit zelfs maar de helft: 3 (i.e. 7,5%) van de 40 alternatieve hypothesen kunnen worden aanvaard. 2 daarvan overlappen met de totale dataset (i.e. value proposition bij representatievoorkeur en internal goal bij efficiëntie in herinneringstaak) en 1 niet (i.e. representatievoorkeur voor competence). Er kan dus gesteld worden dat er vergeleken met het totaal aantal hypothesen relatief weinig alternatieve hypothesen aanvaard konden worden.

Oorzaken hiervan kunnen gevonden worden in zowel het meetinstrument (i.e. experimentele taken) als het construct (i.e. de variabelen). Er zijn vooral aanwijzingen dat dit zou kunnen worden veroorzaakt door het meetinstrument (i.e. experimentele taken). Van de 4 taken lijkt slechts één taak in staat een relevante variabele te meten: de representatievoorkeur taak. In de totale dataset scoren hier immers 2 van de in totaal 5 hypothesen (i.e. 40%) significant, wat ook wordt ondersteund door de kwalitatieve data. Het feit dat beide elkaar ondersteunen is een belangrijke vorm van triangulatie, waardoor aannemelijk is dat deze taak meet wat het moet meten. Bij de overige taken is dit beduidend minder. Bij de notatieassociatie taak is slechts 1 (i.e. interpretatieve effectiviteit) van de 10 hypothesen (10%) significant. Hoewel het resultaat eveneens wordt ondersteund door de kwantitatieve en kwalitatieve representatievoorkeur data is dit slechts het geval bij één van de twee elementen die significant scoorden bij de representatievoorkeur taak, bij de ander scoorde de interpretatieve effectiviteit tijdens de notatieassociatie taak niet significant beter. Daarnaast wordt het resultaat (i.e. interpretatieve effectiviteit) ook niet bevestigd in de begrips- en herinneringstaak, waardoor het effect voor de elementen dus in slechts 1 van de 3 gevallen is waargenomen. Bij de herinneringstaak zijn 2 (i.e. interpretatieve efficiëntie) van de 10 hypothesen (20%) in de totale dataset significant. Beide worden niet ondersteund door de kwantitatieve en kwalitatieve representatievoorkeur data. Ook hier worden ze niet bevestigd door de notatieassociatie- en

begripstaak. Ditzelfde geldt voor de begripstaak, waarbij in de totale dataset 1 (i.e. interpretatieve efficiëntie) van de 10 hypothesen (10%) significant was. Hiermee kan geconcludeerd worden dat de constructvaliditeit van het kwantitatieve en kwalitatieve meetinstrument voor representatievoorkeur als voldoende kan worden bestempeld. Dit is echter niet het geval voor de notatieassociatie-, begrips-, en herinneringstaak. Doordat de notatieassociatie-, begrips-, en herinneringstaak niet goed in staat lijken te zijn om de variabelen (i.e. interpretatieve effectiviteit, -efficiëntie en het construct intuïtieve begripelijkheid) te meten, kan op dit moment niet worden vastgesteld of de onderzoeksresultaten ondersteunend zijn voor de mate waarin de in dit onderzoek gebruikte variabelen in staat zijn om intuïtiviteit te meten. Wel lijken er aanwijzingen te zijn dat dit het geval is voor de representatievoorkeur variabele.

Vervolgonderzoek is nodig om de constructvaliditeit van de representatievoorkeur taak (i.e. kwantitatieve en kwalitatieve uitvraag) te valideren. Daarnaast is het ook noodzakelijk om meetinstrumenten te zoeken die beter in staat zijn om de variabelen interpretatieve effectiviteit en -efficiëntie te meten. Uiteraard is het ook een mogelijkheid om de in dit onderzoek gebruikte meetinstrument fijner af te stellen, omdat deze wellicht niet in staat is om subtiele veranderingen te meten. Andersom is het natuurlijk een mogelijkheid om het gebruikte meetinstrument nogmaals te gebruiken voor meer radicale veranderingen, waardoor beter is uit te sluiten in hoeverre het meetinstrument al dan niet bruikbaar is. Tevens is aandacht nodig voor de mate waarin de variabelen (i.e. interpretatieve effectiviteit en -efficiëntie) in staat zijn om intuïtiviteit te meten. Wanneer de resultaten van de verschillende variabelen elkaar in vervolgonderzoeken niet blijven ondersteunen, dan moeten in de wetenschappelijke literatuur andere variabelen worden gezocht.

5.2.2. Interne validiteit

De interne validiteit heeft van doen met de mate waarin een verondersteld causaal verband daadwerkelijk optreedt. Om verstorende (i.e. derde) factoren in dit onderzoek zoveel mogelijk uit te sluiten is het experiment volledig digitaal aangeboden, zonder menselijke interactie. Hierdoor hebben alle participanten precies dezelfde informatie gekregen en wordt de survey in een beperkte tijdsperiode (i.e. drie weken) gemaakt, waardoor in grote mate sprake zal zijn van dezelfde omstandigheden waaronder het experiment is uitgevoerd. Om te voorkomen dat ongewenste voorkeureffecten de onderzoeksresultaten beïnvloeden, zijn de notaties onderling gewisseld en is gedurende de hele survey dezelfde versie (i.e. combinatie van 'originele' en 'nieuwe' notatie) gebruikt. Alleen in de laatste taak worden beide versies getoond.

Echter is door een deel van de potentiële participanten van het eigen onderzoek - in tegenstelling tot de separate onderzoeken - aangegeven dat de survey als moeilijk werd ervaren. Over de survey werd aangemerkt dat niet altijd duidelijk was wat tijdens de taken van de participant werd verwacht. Daarnaast werd aangegeven dat de modelleertaal niet goed werd begrepen, omdat deze veel jargon bevatte waarmee respondenten niet bekend waren. Dit gegeven lijkt ook te worden bevestigd door het aantal (i.e. 19) deels ingevulde surveys. Daarnaast lijkt dit ook naar voren te komen bij de verschillen in modelleerexpertise en -ervaring tussen de participanten in de eigen dataset en die van de separate onderzoeken: participanten van het eigen onderzoek lijken aanzienlijk vaker geen enkele modelleertaal ervaring te hebben; minder ervaring te hebben met BPM-modelleertalen; en vaker geen enkele modelleerexpertise te hebben ten opzichte van participanten in de andere datasets. Ook is er een aanzienlijk verschil zichtbaar in het aantal hypothesen die konden worden aanvaard van de totale (i.e. 6) en eigen (i.e. 3) dataset. Daarnaast wordt de significant geteste representatievoorkeur

voor element competence niet bevestigd door de kwalitatieve data. Hiermee lijkt het erop dat de specifieke eigenschappen van de participanten van het eigen onderzoek een verstoring effect hebben gehad op de onderzoeksresultaten. Concluderend is enige voorzichtigheid geboden omtrent de interne validiteit. Het is dan ook aan te raden om in vervolgonderzoek vooraf goed na te denken over de benodigde modelleerexpertise en -ervaring van de participanten, zodat dit geen verstoring effect kan hebben op de onderzoeksresultaten.

5.2.3. Externe validiteit

De externe validiteit gaat over in hoeverre onderzoeksresultaten gegeneraliseerd kunnen worden. In dit onderzoek is hiermee rekening gehouden door gebruik te maken van participanten die meer overeenkomen met de daadwerkelijke eindgebruikers van de PGA-notatie (i.e. business users in een praktische bedrijfscontext) dan de participanten uit voorgaand onderzoek (i.e. masterstudenten). Hierdoor zouden de resultaten beter te generaliseren zijn dan die uit voorgaande onderzoeken.

Echter, zoals in vorig onderdeel al werd genoemd zijn er tussen de participanten onderling verschillen (i.e. modelleerervaring en -expertise) te zien die ook een effect lijken te hebben gehad op de onderzoeksresultaten: participanten van de eigen dataset beoordelen de nieuwe notatie minder vaak (i.e. 3 keer) als significant beter in vergelijking met de totale dataset (i.e. 6 keer). Tussen de eigen en andere datasets zijn de volgende verschillen te zien: in de eigen dataset hebben participanten aanzienlijk vaker geen enkele modelleertaal ervaring; minder ervaring met BPM-modelleertalen; en vaker geen enkele modelleerexpertise ten opzichte van participanten in de andere datasets. Als er met dergelijke verschillen in participanteigenschappen al dermate grote verschillen in de onderzoeksresultaten zichtbaar zijn, dan kan de externe validiteit als twijfelachtig worden bestempeld. De resultaten van het onderzoek zijn hiertoe alleen te veralgemenen voor het vergelijken van de PGA-notatie door business users (i.e. voornamelijk mannen van 45 jaar met master- of bachelor graad) met een gemiddelde modelleerervaring in de praktische bedrijfscontext (i.e. Nederlandse organisaties in de private en publieke sector). Over het algemeen zijn deze bekend met conceptuele modelleertalen voor Business Process Management, Requirements Engineering en Enterprise Architecture, maar soms ook met geen enkele modelleertaal. In de afgelopen 5 jaar hebben zij gemiddeld tussen de 11 en 20 conceptuele modellen gelezen en tussen de 1 en 10 gemaakt.

In vervolgonderzoek is het aan te raden om vooraf goed stil te staan bij de modelleerervaring en -expertise eigenschappen die participanten zouden moeten hebben in relatie tot de modelleertaal die wordt geëvalueerd, wat naar verwachting zorgt voor meer robuuste resultaten die daardoor beter zijn te veralgemenen. Anders is de kans groot dat niet de intuïtieve begrijpelijkheid wordt gemeten, maar de mate waarin modelleerervaring en -expertise nodig is om de modelleertaal te begrijpen.

5.2.4. Betrouwbaarheid

Bij betrouwbaarheid wordt gekeken naar tot welke hoogte onderzoeksresultaten reproduceerbaar zijn. Dit is in het onderzoek geborgd door de dataverzameling conform het in hoofdstuk 3 beschreven plan plaats te laten vinden. Zo zijn er specifieke participanten (i.e. business users in een praktische bedrijfscontext) benaderd en is het onderzoek online uitgevoerd. Dit heeft als voordeel dat alle participanten tijdens het invullen van de survey dezelfde informatie hebben gekregen. Daarnaast gebeurt de dataverzameling digitaal, waardoor bij de vastlegging geen sprake is van eventuele subjectiviteit van de onderzoeker. Vervolgens is de data op een grondige manier geanalyseerd en zijn

de resultaten uit SPSS en in excel vastgelegd, zodat de uitkomsten herleidbaar en reproduceerbaar zijn, zie bijlage 8 tot en met 12. Bij de analyse is ook gebruik gemaakt van een combinatie van objectieve (i.e. interpretatieve effectiviteit en interpretatieve efficiëntie) en subjectieve (i.e. representatievoorkeur) variabelen alsook kwalitatieve data (i.e. representatievoorkeur), waardoor triangulatie optreedt zodat resultaten als meer betrouwbaar kunnen worden gezien. Echter, voor een deel hebben participanten aan het onderzoek meegedaan die onderling van elkaar afwijken met betrekking tot de modelleerervaring en -expertise, iets wat invloed lijkt te hebben gehad op de onderzoeksresultaten. Het aantal significante resultaten, alsook de hypothesen waarbij dit gebeurde verschillen immers van elkaar. Aangezien de drie separate onderzoeken al niet hebben geleid tot dezelfde onderzoeksresultaten, kan niet zondermeer worden verwacht dat het repliceren van het onderzoek op een ander moment wel zal leiden tot dezelfde resultaten. Concluderend kan de betrouwbaarheid van de evaluatieresultaten als matig worden bestempeld.

5.2.5. Ethische Aspecten

Voor wat betreft de ethische aspecten wordt in ogenschouw genomen in hoeverre het gedrag van de onderzoeker verantwoord is in relatie tot diegenen die onderwerp worden van het onderzoek, of die daardoor worden beïnvloed (Saunders et al., 2019). In dit onderzoek is hiermee rekening gehouden door participanten bij de start van de survey digitaal een instemmingsformulier voor deelname aan het onderzoek (i.e. declaration of consent) te laten tekenen, waarmee participanten aangeven op de hoogte zijn van het feit dat het een wetenschappelijk onderzoek betreft; dat deelname vrijwillig is en te allen tijde kan worden beëindigd; en dat data anoniem zal worden verzameld en daartoe niet herleidbaar is naar de participant. Deze beloften zijn ook technisch gerealiseerd: zo is voorzien in de mogelijkheid dat de participant tijdens het experiment op elk willekeurig moment kan stoppen; en wordt de verzamelde data tijdens het invullen automatisch geanonimiseerd. Tenslotte is de data tijdens de analyse vertrouwelijk behandeld en zijn participantnamen alsook de bedrijfsnaam van de casusorganisatie geanonimiseerd. Concluderend is in voldoende mate rekening gehouden met ethische aspecten.

5.3. Conclusies

Dit hoofdstuk bevat de belangrijkste conclusies op basis van de discussie en reflectie, waarmee een antwoord wordt gegeven op de vraag in hoeverre de 'nieuwe' PGA-notatie intuïtiever is dan de 'originele'. Daarnaast zal ook een uitspraak worden gedaan over de mate waarin de evaluatietechniek van Bork et al. (2019) heeft geleid tot significante verbeteringen in de intuïtiviteit van de PGA-notatie voor eindgebruikers in een praktische bedrijfscontext. Eveneens zal een besluit worden genomen over eventueel benodigde aanpassingen aan de experimentele opzet van dit onderzoek, waarmee tevens ook een antwoord wordt gegeven op de centrale onderzoeksvraag *'Hoe kan de intuïtiviteit van verschillende versies van een DSML-notatie empirisch met elkaar worden vergeleken in een praktische bedrijfscontext?'*.

In hoeverre de nieuwe PGA-notatie intuïtiever is dan originele

Op basis van de onderzoeksresultaten en de hierop volgende reflectie kan geconcludeerd worden dat bij 2 van de 5 elementen de 'nieuwe' notaties intuïtiever zijn dan de 'originele': value proposition en customer goal. Op basis van de validiteit reflectie kan immers geconcludeerd worden dat alleen de constructvaliditeit van het kwantitatieve en kwalitatieve meetinstrument voor representatievoorkeur als voldoende kan worden bestempeld, aangezien de statistische tests eveneens worden ondersteund door de kwalitatieve data. Dit gegeven is een belangrijke vorm van triangulatie, waardoor aannemelijk is dat de representatievoorkeur taak meet wat het moet meten. Dit is echter niet het geval voor de notatieassociatie-, begrips-, en herinneringstaak, waardoor geconcludeerd kan worden dat deze niet goed in staat lijken te zijn om de variabelen interpretatieve effectiviteit, -efficiëntie en het construct intuïtieve begrijpelijkheid te meten. Onduidelijk is nog of gezocht moet worden naar andere meetinstrumenten die beter in staat zijn om de variabelen te meten, of dat deze fijner afgesteld moeten worden. Wellicht dat de gebruikte instrumenten niet in staat zijn om subtiele verbeteringen te meten maar wel meer radicale. Aangezien de variabelen interpretatieve effectiviteit, -efficiëntie niet goed gemeten konden worden kan op dit moment dan ook niet worden bevestigd of deze variabelen in staat zijn om intuïtiviteit te meten. Wel lijken er dus aanwijzingen te zijn dat dit het geval is voor de representatievoorkeur variabele.

Ook is enige voorzichtigheid is geboden omtrent de interne validiteit. Het lijkt er namelijk op dat onderlinge verschillen tussen participanten (i.e. modelleerexpertise en -ervaring) een verstorend effect hebben gehad op de onderzoeksresultaten. Door een deel van de potentiële participanten van het eigen onderzoek - in tegenstelling tot de separate onderzoeken - is namelijk aangegeven dat de survey als moeilijk werd ervaren. In de demografische data zijn dan ook verschillen te zien tussen de eigen dataset en die van de separate onderzoeken: participanten van het eigen onderzoek lijken aanzienlijk vaker geen enkele modelleertaal ervaring te hebben; minder ervaring te hebben met BPM-modelleertalen; en vaker geen enkele modelleerexpertise te hebben ten opzichte van participanten in de andere datasets. Ook zijn er verschillen te zien in de onderzoeksresultaten: participanten van de eigen dataset beoordelen de nieuwe notatie minder vaak (i.e. 3 keer) als significant beter in vergelijking met de totale dataset (i.e. 6 keer). Dit heeft eveneens consequenties voor de externe validiteit: hoewel participanten in dit onderzoek meer overeenkomen met de daadwerkelijke eindgebruikers van de PGA-notatie (i.e. business users in een praktische bedrijfscontext) dan de participanten uit voorgaand onderzoek (i.e. masterstudenten), zijn resultaten door deze onderlinge verschillen toch niet beter te generaliseren. Als er namelijk met dergelijke verschillen in participanteneigenschappen al dermate grote verschillen in de onderzoeksresultaten zichtbaar zijn, dan kan ook de externe validiteit als twijfelachtig worden bestempeld. Ditzelfde kan worden geconcludeerd over de betrouwbaarheid: de drie separate onderzoeken hebben immers al niet tot

dezelfde onderzoeksresultaten geleid. Hierdoor kan niet zondermeer worden verwacht dat het repliceren van het onderzoek zal leiden tot dezelfde resultaten. Concluderend kan de betrouwbaarheid van de evaluatieresultaten als matig worden bestempeld.

Mate waarin evaluatietechniek Bork heeft geleid tot significante verbeteringen PGA-notatie

De experimentele evaluatietechniek van Bork et al. (2019) heeft tot verbeteringen in 5 elementen van de PGA-notatie geleid. In dit onderzoek is naar voren gekomen dat voor 2 (i.e. 40%) van die elementen (i.e. value proposition en customer goal) de nieuwe notatie als significant beter is beoordeeld door business users in een praktische bedrijfscontext in vergelijking met de originele. De onderzoeksresultaten laten immers zien dat de kwantitatieve en kwalitatieve representatievoorkeur data hierop wijzen. Echter, dit is niet vastgesteld voor de variabelen interpretatieve effectiviteit, - efficiëntie en het totale construct intuïtieve begrijpelijkheid, die zijn gemeten met behulp van de meetinstrumenten notatieassociatie-, begrips-, en herinneringstaak. Dit lijkt een opvallend gegeven, aangezien de evaluatietechniek van Bork et al. (2019) verbeteringen heeft geformuleerd juist op basis van deels dezelfde taken: de notatieassociatie- en begripstaak. Hierdoor lijkt het logisch dat nieuwe notaties beter op deze taken zouden moeten scoren dan de originele. Hoewel kanttekeningen geplaatst kunnen worden bij de validiteit en betrouwbaarheid van de resultaten van dit onderzoek, is het de verwachting dat het gebruikte meetinstrument (i.e. notatieassociatie-, begrips-, en herinneringstaak) geen subtiele maar wel meer radicale verbeteringen kan detecteren. Het gegeven dat dit bij 60% van de nieuwe notaties niet het geval is, leidt tot de conclusie dat de evaluatietechniek van Bork et al. (2019) over het algemeen heeft geleid tot subtiele verbeteringen die door eindgebruikers deels als significant beter worden beoordeeld.

Hoe intuïtiviteit van DSML-notaties empirisch kunnen worden vergeleken in een praktische bedrijfscontext

Voorlopig kan worden aangenomen dat de intuïtiviteit van verschillende versies van een DSML-notatie empirisch met elkaar kan worden vergeleken in een praktische bedrijfscontext, door gebruik te maken van de representatievoorkeur variabele. Deze variabele kan worden gemeten met een online aangeboden representatievoorkeur taak waarin zowel kwantitatieve als kwalitatieve data wordt verzameld, waardoor vervolgens middels triangulatie kan worden vastgesteld in welke mate de kwalitatieve data de resultaten van de statistische tests op de kwantitatieve data ondersteunen. Niet geschikt voor het detecteren van subtiele verbeteringen - maar wellicht wel voor meer radicale verbeteringen - zijn de in het onderzoek gebruikte meetinstrumenten voor interpretatieve effectiviteit en -efficiëntie: notatieassociatie-, begrips- en herinneringstaak. Ook is niet zeker of de variabelen interpretatieve effectiviteit en -efficiëntie wel geschikt zijn voor het meten van intuïtieve begrijpelijkheid. Tevens lijkt de vergelijking van DSML-notaties niet te kunnen worden gemaakt wanneer de modelleerexpertise en -ervaring van participanten onderling te veel verschillen.

Op basis van bovenstaande kan geconcludeerd worden dat het onderzoek een bijdrage levert aan de body of knowledge over het empirisch evalueren van de intuïtiviteit van DSML-notaties. Hoewel voor verbetering vatbaar levert het namelijk een startpunt voor een methode die de intuïtieve begrijpelijkheid van verschillende versies van een DSML-notatie empirisch met elkaar vergelijkt. Ook is het onderzoek - in tegenstelling tot voorgaand onderzoek - gedaan in een praktische bedrijfscontext, waardoor inzichtelijk is geworden dat mogelijk extra aandacht nodig is bij het selecteren van participanten (i.e. modelleerexpertise en -ervaring) voor het evalueren van DSML's. De laatste bijdrage heeft van doen met het feit dat een eerste inzicht is verkregen in hoezeer de evaluatietechniek van Bork et al. (2019) tot een verbeterde DSML-notatie heeft geleid.

5.4. Aanbevelingen voor de praktijk

Zoals eerder beschreven wordt intuïtiviteit gekoppeld aan een verbeterde klantervaring en heeft dus een positief effect op de gebruiksvriendelijkheid en acceptatie van DSML's. Hiermee is het dus in het belang van de organisaties achter de DSML's (e.g. BiZZDesign) om deze intuïtief te maken en hierop periodiek verbeteringen door te voeren, waardoor de notatie steeds intuïtiever wordt. Immers, hoe beter een DSML wordt beoordeeld door eindgebruikers, hoe groter de kans op succes voor de organisaties achter de DSML's. Om hierbij te helpen is het allereerst van belang om een evaluatietechniek te hebben waarmee de intuïtiviteit van notaties kan worden vastgesteld. Hier hebben Bork et al. (2019) reeds een aanzet toe gedaan. Vervolgens is het van belang om te toetsen of potentiële verbeteringen op de notatie daadwerkelijk als intuïtiever worden beoordeeld. Het is tenslotte weinig zinvol om veranderingen door te voeren op een modelleertaal, die de intuïtiviteit van notaties niet verbetert en wellicht zelfs verslechtert. Hiertoe heeft dit onderzoek een eerste aanzet geleverd voor een vervolgevaluatie die - na verder te zijn fijn geslepen - in de toekomst mogelijk aanvullend op de evaluatietechniek van Bork et al. (2019) in de praktijk kan worden ingezet voor het vergelijkend evalueren van verschillende versies van een DSML. Met een dergelijk instrument kunnen potentiële verbeteringen in DSML's worden getest en besluiten hierover beter worden onderbouwd. Wellicht dat het DSML-ontwerpers ook stimuleert om vaker en betere verbeteringen door te voeren, omdat empirisch bewezen kan worden in welke mate ze beter zijn dan de oude.

In de praktijk kan de evaluatie in de toekomst als volgt worden opgezet: organisaties kunnen een meer fijn geslepen versie van de evaluatietechniek van Bork et al. (2019) inzetten bij een groep participanten die overeenkomen met de daadwerkelijke gebruikers van de DSML. Aangezien intuïtiviteit gaat over het gemak waarmee een notatie direct kan worden begrepen, zonder enige voorkennis of training, is het namelijk aan te raden om tijdens de evaluatie te kiezen voor participanten die de modelleertaal nog niet kennen. Hierbij is het wel belangrijk om goed te onderzoeken wat de specifieke eigenschappen van de daadwerkelijke eindgebruikers zijn, met name omtrent de modelleerexpertise en -ervaring. Dit inzicht kan vervolgens worden gebruikt om geschikte participanten te selecteren. De verbeteringen die voortkomen uit de evaluatietechniek van Bork et al. (2019) kunnen vervolgens met een verbeterde versie van de vervolgevaluatie worden beoordeeld door een andere groep participanten die overeenkomen met de eindgebruikers. Om er nadien zeker van te zijn dat de verbeteringen eveneens als beter worden beoordeeld door de bestaande gebruikers is het van belang om daadwerkelijke eindgebruikers de potentiële verbeteringen - op basis van de vervolgevaluatie met participanten die de modelleertaal nog niet kennen - te laten valideren. Bij de aanvang van de ingebruikname van de evaluatie- en vervolgevaluatietechniek zou het logisch zijn dat organisaties namelijk eerst vertrouwen moeten krijgen over de mate waarin de resultaten representatief zijn voor de daadwerkelijke eindgebruikers. Dit zijn immers de klanten van de organisatie die organisaties uiteraard niet willen verliezen. Wanneer voldoende vertrouwen is opgebouwd met de vervolgevaluatie, dan is deze extra stap wellicht niet meer nodig.

5.5. Aanbevelingen voor verder onderzoek

Aanbevelingen voor vervolgonderzoek zijn vooral gericht op het verbeteren van de construct-, interne en externe validiteit alsook de betrouwbaarheid.

Op basis van de onderzoeksresultaten is voorlopig aangenomen dat de variabele representatievoorkeur de intuïtieve begrijpelijkheid meet en dat dit kan middels de kwantitatieve en kwalitatieve resultaten van de representatievoorkeur taak. Echter, om hier zeker van te kunnen zijn zal in een vervolgonderzoek deze taak moeten worden herhaald om de constructvaliditeit van de representatievoorkeur taak te kunnen valideren.

Ook is voorlopig aangenomen dat de meetinstrumenten notatieassociatie-, begrips- en herinneringstaak niet goed in staat zijn om de variabelen interpretatieve effectiviteit en -efficiëntie te meten. Dit heeft mogelijk ook consequenties voor evaluatietechniek van Bork et al. (2019) aangezien deze deels gebruik maakt van dezelfde taken (i.e. de notatieassociatie- en begripstaak) en slechts lijkt te hebben geleid tot subtiele verbeteringen die door eindgebruikers deels als significant beter worden beoordeeld. Daarom is het ook noodzakelijk om in vervolgonderzoek meetinstrumenten te zoeken die beter in staat zijn om de variabelen interpretatieve effectiviteit en -efficiëntie te meten. Wellicht dat de papers uit het theoretisch kader hiervoor een startpunt kunnen zijn. Uiteraard is het ook een mogelijkheid om de meetinstrumenten in vervolgonderzoek nogmaals te gebruiken voor het evalueren van meer radicale veranderingen, waardoor beter is uit te sluiten in hoeverre de meetinstrumenten al dan niet bruikbaar zijn. Andersom geldt hetzelfde: zo is het ook een mogelijkheid om de in dit onderzoek gebruikte meetinstrumenten fijner af te stellen, omdat deze op dit moment misschien niet in staat zijn om subtiele verbeteringen te meten. Zo kunnen bij de notatieassociatie taak per element meer vragen worden gesteld over de intuïtiviteit van specifieke kenmerken van een notatie. Bijvoorbeeld bij de nieuwe notatie van internal goal kan gevraagd worden naar de associatie met het ronde kader, waardoor vastgesteld kan worden in hoeverre deze geassocieerd wordt met de beoogde betekenis goal. Vervolgens kan ook naar de associatie worden gevraagd van het fabriek symbool met daarin de tandwielen, waardoor duidelijk wordt in hoeverre deze wordt geassocieerd met de beoogde betekenis internal. Op die manier is beter vast te stellen welk onderdeel van een notatie al dan niet als intuïtiever wordt beoordeeld. Ook kan de taak worden aangevuld met de mogelijkheid voor participanten om bij het gegeven antwoord een toelichting te geven, waarmee ook inzicht wordt verkregen in hoeverre de participant de juiste associatie maakt tussen het element en de betekenis. De mogelijkheid bestaat namelijk dat de participant wel de juiste associatie maakt, maar het verkeerde woord gebruikt om de betekenis van het element te duiden. Ook voor de begripsvragen kunnen meer vragen worden gesteld per element, waardoor een meer genuanceerd beeld kan ontstaan over hoe de intuïtiviteit wordt beoordeeld. Hierbij kan een mix gemaakt worden tussen eenvoudigere en moeilijkere vragen om een beter beeld te krijgen bij hoe goed de elementen worden begrepen. Uiteraard moeten deze vragen worden gemixt om eventuele leereffecten te voorkomen. Eveneens in deze taak kan participanten de mogelijkheid worden geboden om het gegeven antwoord toe te lichten, waarmee de kwantitatieve resultaten van de begripstaak kunnen worden ondersteund. Voor wat betreft de herinnertaak kan ervoor worden gekozen om deze tijdens het experiment twee keer uit te voeren, op twee verschillende momenten in het onderzoek (e.g. op oorspronkelijke plek en na de representatievoorkeur taak). Hiermee zou een meer genuanceerd beeld kunnen ontstaan over de mate waarin de betekenis van het element wordt onthouden naarmate de participant de betekenis vaker heeft gezien.

In dit onderzoek was het niet mogelijk om een uitspraak te doen over de mate waarin de variabelen interpretatieve effectiviteit en -efficiëntie in staat zijn om intuïtiviteit te meten. Vandaar dat hier in vervolgonderzoek ook extra aandacht aan moet worden besteed. Wanneer de resultaten van de verschillende variabelen (i.e. interpretatieve effectiviteit, -efficiëntie en representatievoorkeur) elkaar in vervolgonderzoeken namelijk wederom niet ondersteunen, dan moet in de wetenschappelijke literatuur wellicht naar andere variabelen worden gezocht die wel in staat zijn om de intuïtiviteit van DSML's te meten.

Bij de reflectie op interne validiteit is naar voren gekomen dat verschillen (i.e. modelleerexpertise- en ervaring) tussen participanten mogelijk een effect hebben gehad op de onderzoeksresultaten. Daarom is het aan te raden om in vervolgonderzoek vooraf goed stil te staan bij de modelleerervaring en -expertise eigenschappen die participanten zouden moeten hebben in relatie tot de modelleertaal die wordt geëvalueerd. Deze eigenschappen zouden vooraf vastgesteld moeten worden door onderzoek hiernaar te doen bij daadwerkelijke eindgebruikers. Vervolgens kunnen deze eigenschappen worden gebruikt als inclusie-/exclusie criteria bij het selecteren van participanten. Het gebruiken van dezelfde type participanten zal er tevens toe leiden dat bij het repliceren van het onderzoek dezelfde resultaten worden verkregen. Concluderend zal voorgaande een positief effect hebben op de interne en externe validiteit, alsook de betrouwbaarheid, waardoor de onderzoeksresultaten ook beter zijn te veralgemenen.

Referenties

- Aranda, J., Ernst, N., Horkoff, J., & Easterbrook, S. (2007). *A framework for empirical evaluation of model comprehensibility*. Paper presented at the International Workshop on Modeling in Software Engineering (MISE'07: ICSE Workshop 2007).
- BiZZdesign. (2018). Enterprise Studio - The leading modeling tool with native ArchiMate support. Retrieved from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=10&cad=rja&uact=8&ved=2ahUKEwix8JDfoK7oAhUCuqQKHXYIDLEQFjAJegQIBxAB&url=https%3A%2F%2Fbizzdesign.com%2Fproducts%2Fenterprise-studio%2F&usg=AOvVaw37r5IEOt3oPd-C97_sappM
- Bork, D., Karagiannis, D., & Pittl, B. (2018). *Systematic analysis and evaluation of visual conceptual modeling language notations*. Paper presented at the 2018 12th International Conference on Research Challenges in Information Science (RCIS).
- Bork, D., Karagiannis, D., & Schrüffer, C. (2019). *Intuitive Understanding of Domain-specific Modeling Languages: Proposition and Application of an Evaluation Technique*. Paper presented at the International Conference on Conceptual Modeling.
- Cook, C., Heath, F., Thompson, R. L., & Thompson, B. (2001). Score reliability in webor internet-based surveys: Unnumbered graphic rating scales versus likert-type scales. *Educational and Psychological Measurement*, 61(4), 697-706.
- Figl, K. (2017). Comprehension of procedural visual business process models. *Business & Information Systems Engineering*, 59(1), 41-67.
- Figl, K., Mendling, J., & Strembeck, M. (2013). The influence of notational deficiencies on process model comprehension. *Journal of the Association for Information Systems*, 14(6), 1.
- Figl, K., & Recker, J. (2016). Exploring cognitive style and task-specific preferences for process representations. *Requirements Engineering*, 21(1), 63-85.
- Frank, U. (2010). Outline of a Method for Designing Domain-Specific Modelling Languages. Retrieved from <https://www.econstor.eu/obitstream/10419/58163/1/716089785.pdf>
- Gartner. (2019). Gartner's Cool Vendors in AI for Banking and Investment Services. Retrieved from <https://blogs.gartner.com/moutusi-sau/2019/05/14/gartners-cool-vendors-ai-banking-investment-services/>
- Gemino, A., & Wand, Y. (2004). A framework for empirical evaluation of conceptual modeling techniques. *Requirements Engineering*, 9(4), 248-260.
- Group, T. O. (2019). ArchiMate® 3.1 Specification. Retrieved from <https://pubs.opengroup.org/architecture/archimate3-doc/>
- Jošt, G., Huber, J., Heričko, M., & Polančič, G. (2016). An empirical investigation of intuitive understandability of process diagrams. *Computer Standards & Interfaces*, 48, 90-111.
- Lankhorst, M. (2009). *Enterprise Architecture at Work: Modelling, Communication and Analysis* (Second edition ed.).
- McKinsey. (2017). McKinsey on Digital Services - Introducing the next-generation operating model. Retrieved from <https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Introducing%20the%20next-generation%20operating%20model/Introducing-the-next-gen-operating-model.ashx>
- Moody, D. (2009). The "physics" of notations: toward a scientific basis for constructing visual notations in software engineering. *IEEE Transactions on software engineering*, 35(6), 756-779.
- Moody, D., Heymans, P., & Matulevičius, R. (2010). Visual syntax does matter: improving the cognitive effectiveness of the i* visual notation. *Requirements Engineering*, 15(2), 141-175.
- Roelens, B., & Bork, D. (2020). Performing an experimental evaluation to improve the initial PGA notation.

- Roelens, B., Steenacker, W., & Poels, G. (2019). Realizing strategic fit within the business architecture: the design of a Process-Goal Alignment modeling and analysis technique. *Softw Syst Model*, 18, 631–662. Retrieved from <https://doi.org/10.1007/s10270-016-0574-5>
- Rolon, E., Sanchez, L., Garcia, F., Ruiz, F., Piattini, M., Caivano, D., & Visaggio, G. (2009). *Prediction models for BPMN usability and maintainability*. Paper presented at the 2009 IEEE Conference on Commerce and Enterprise Computing.
- Saunders, M., Lewis, P., & Thornhill, A. (2019). *Research methods for business students*: Pearson education.
- Studiedata, S. H. (2020). Toetsmatrix. Retrieved from <https://sh-studiedata.nl/R/Toetsmatrix.html>

Bijlage 1: Begrippenlijst

Tabel 21: Begrippenlijst 1/2

Begrip	Betekenis
BCM	Business Continuity Management, een DSML
Begripstaak	Experimentele taak die doorgaans bestaat uit een reeks vragen over notatie-elementen waarbij de notatie wordt getoond
Betrouwbaarheid	Tot welke hoogte resultaten reproduceerbaar zijn
Between-subjects design	Experimenteel design met zowel een experimentele groep als een controlegroep, waarbij alleen de experimentele groep de interventie ondergaat. Door deze opzet zijn bij dit type design de onderzoeksresultaten tussen de experimentele en controlegroep met elkaar te vergelijken, waarbij het voor de hand ligt dat eventuele verschillen verklaard kunnen worden door de interventie
BPMN	Business Process Modeling and Notation, een GPML
Construct intuïtieve begripelijkheid	Geaggregeerde score van interpretatieve effectiviteit, -efficiëntie en representatievoorkeur
Constructvaliditeit	Tot welke hoogte het meetinstrument meet wat het moet meten
Daadwerkelijke eindgebruikers PGA-notatie	Business users in een praktische bedrijfscontext
Deep-level understanding	Mate waarin een persoon de notatie toe kan passen
DSML	Domain-Specific Modeling Languages: domeinspecifieke EA-modelleertalen die bedoeld zijn voor bepaalde toepassingsdomeinen en modelleerconcepten bevatten die verrijkt zijn met termen en visualisaties waarmee eindgebruikers bekend zijn
EA	Enterprise Architectuur: een samenhangend geheel van principes, methoden en modellen die worden gebruikt bij het ontwerpen en realiseren van de organisatiestructuur, bedrijfsprocessen, informatiesystemen en infrastructuur van een organisatie
EA-modelleertalen	Conceptuele modellen die voor begrips- en communicatiedoelinden beogen om de realiteit van een specifiek aandachtsgebied minder complex weer te geven, door alleen de hiervoor relevante aspecten te benadrukken
Externe validiteit	In hoeverre onderzoeksresultaten gegeneraliseerd kunnen worden
GPML	General-Purpose Modeling Languages: generieke EA-modelleertalen die onafhankelijk zijn van het gebruiksdomein en daarom geschikt zijn voor een breed scala aan domeinen
Herinneringstaak	Experimentele taak waarin vragen worden gesteld over de notatie, nadat deze is weggenomen
Interne validiteit	Mate waarin een verondersteld causaal verband daadwerkelijk optreedt
Interpretatieve effectiviteit	Een afhankelijke variabele in het onderzoek, waarbij wordt gekeken naar het aantal correcte antwoorden op de notatieassociatie-, begrips-, herinneringstaken
Interpretatieve efficiëntie	Een afhankelijke variabele in het onderzoek, waarbij wordt gekeken naar de tijd in seconden die de participant nodig heeft om de notatieassociatie-, begrips-, herinneringstaken af te ronden
Intuïtiviteit/intuïtieve begripelijkheid	Het gemak waarmee een notatie direct kan worden begrepen door gebruikers, zonder enige voorkennis of training

Nieuwe notatie	PGA-element die middels de verbetertechniek van van Bork et al. (2019) is verbeterd en hiertoe heeft geleid tot een nieuwe notatie
----------------	--

Tabel 22: Begrippenlijst 2/2

Begrip	Betekenis
Notatie	Grafische visualisatie van modelleerelementen en relaties, onafhankelijke variabele in het onderzoek
Originele notatie	Oorspronkelijk PGA-element, voordat deze middels de verbetertechniek van Bork et al. (2019) is verbeterd
Perceptual discriminability	Mate waarin symbolen onderling duidelijk van elkaar onderscheiden kunnen worden
PGA	Process-Goal Alignment: een DSML met als doel om strategische fit te bereiken in de bedrijfsarchitectuur en met een heat map inzichtelijk te maken in welke mate bedrijfsactiviteiten (i.e. interne infrastructuur en processen) de bedrijfsstrategie ondersteunen
Representatievoorkeur	Een afhankelijke variabele in het onderzoek, waarbij wordt gekeken naar welke notatie (i.e. 'nieuwe' of 'originele' versie) de intuïtieve voorkeur uitgaat bij het zien van beide notaties in de representatievoorkeur taken
Representatievoorkeur taak	Experimentele taak waarin wordt gekeken naar welke versie (i.e. 'nieuw' of 'origineel') van de PGA-notatie de intuïtieve voorkeur uitgaat bij het zien van beide notaties
Semantic transparency	Mate waarin symbolen hun betekenis suggereren
Semiotic clarity	Mate waarin semantische constructen en grafische symbolen 1-op-1 met elkaar overeenkomen
Strategic fit	Mate waarin bedrijfsactiviteiten (i.e. interne infrastructuur en processen) de bedrijfsstrategie ondersteunen
Surface-level understanding	Mate waarin een persoon een notatie begrijpt
UML	Unified Modeling Language, een GPML
Within-subjects design	Experimenteel design waarbij het experiment wordt uitgevoerd binnen één groep, waarbij alle participanten worden blootgesteld aan de geplande interventie

Bijlage 2: Gedetailleerde analyse paper selectie theoretisch kader

Onderstaand de gedetailleerde analyse die heeft geleid tot de zes papers die gebruikt zijn voor het opzetten van het theoretisch kader. Tabel 23 bevat de resultaten van de query, tabel 24 bevat de resultaten van de forward snowballing op Figl et al. (2013). Voor de volledigheid zijn in tabel 25 en 26 ook de forward snowballing resultaten toegevoegd op de papers van respectievelijk Moody et al. (2010) en Figl (2017). In de tabellen zijn de papers die een link hebben met intuïtiviteit gemarkeerd met donkergroen, de papers die relevant zijn voor het onderwerp - maar geen link hebben met intuïtiviteit - met lichtgroen. Papers die een link hebben met intuïtiviteit en in zowel de query als de forward snowballing naar voren zijn gekomen zijn geel gemarkeerd.

In principe zijn alle vijf de papers met een link naar intuïtiviteit meegenomen in het theoretisch kader. Van de papers die voor het onderwerp relevant zijn - zonder link met intuïtiviteit - is de paper van Aranda et al. (2007) gekozen, omdat de titel trefwoorden (i.e. empirical evaluation en comprehensibility) bevat die het meest relevant zijn voor het beantwoorden van de onderzoeksvraag. Empirical evaluation is van belang, omdat dit trefwoord onderdeel is van de onderzoeksvraag. Comprehensibility is belangrijk, omdat deze term nauw verwant is aan intuïtiviteit, eveneens onderdeel van de onderzoeksvraag.

Tabel 19: Analyseresultaten building blocks query

Referentie paper	Titel	Auteur	Jaar van publicatie	Aantal referenties in Web of Science	Methodologie paper, individueel experiment, niet van toepassing (M;IE;nvt)	Doel: link met intuïtiviteit? (J;N)	Typen vragen: recall, comprehension, problem-solving, cloze (R;CO;PS;CL)	Soort respondenten: students vs business experts (ST;BE)
1	Guidelines for using UML association classes and their effect on domain understanding in requirements engineering	Bera, Palash; Bera, Palash; Evermann, Joerg	2014	9	IE	N	PS	ST
2	Improving the representation of roles in conceptual modelling: theory, method, and evidence	Bera, Palash; Bera, Palash; Burton-Jones, Andrew	2017	0	IE	N	PS	ST
3	An automated framework for detection and resolution of cross references in legal texts	Sannier, N; Adedjouma, M; Sabetzadeh, M	2017	3	nvt			
4	The Influence of Notational Deficiencies on Process Model Comprehension	Figl, K; Mendling, J; Strembeck, M	2013	31	IE	J	CO;PS	ST
5	Comprehension of Procedural Visual Business Process Models	Figl, Kathrin	2017	19	M	J	R;CO;PS;CL	
6	In Search of What We Experimentally Know about Unit Testing	Juristo, N; Moreno, A.M; Vegas, S	2006	24	nvt			
7	A Framework for Empirical Evaluation of Model Comprehensibility	Aranda, Jorge; Ernst, Neil; Horkoff, Jennifer	2007	0	M	N	CO	
8	Similarity measures in formal concept analysis	Alqadah, Faris; Bhatnagar, Raj	2011	23	nvt			
9	Understanding Declare models: strategies, pitfalls, empirical results	Haisjackl, Cornelia; Haisjackl, Cornelia; Barba, Irene	2016	0	IE	N	CO;PS	ST
10	A systematic review of transformation approaches between user requirements and analysis models	Yue, Tao; Yue, Tao; Briand, Lionel C	2011	0	nvt			
11	Provision of decision support in environmental management	Piirimäe, Kristjan	2011	0	nvt			
12	RiskM: A multi-perspective modeling method for IT risk assessment	Strecker, Stefan; Strecker, Stefan; Heise, David	2011	23	nvt			
13	Visual syntax does matter: Improving the cognitive effectiveness of the i visual notation	Moody, D.L; Heymans, Patrick; Matulevicius, Raimundas	2010	51	M	J	R;CO;PS	
14	An evaluative baseline for geo-semantic relatedness and similarity	Ballatore, Andrea; Bertolotto, Michela; Wilson, David C	2014	15	nvt			
15	Effect of personality type on structured tool comprehension performance	Gorla, Narasimhaiah; Gorla, Narasimhaiah; Chiravuri, Ananth	2013	1	nvt			
16	Simplification and aggregation strategies applied for factory analysis in conceptual phase using simulation	Moris, Matias; Ng, Amos; Svensson, Jacob	2008	0	nvt			
17	An Approach for Integrating the Prioritization of Functional and Nonfunctional Requirements	Dabbagh, M; Lee, SP	2014	8	nvt			

Tabel 20: Analyseresultaten forward snowballing Figl et al. (2013)

Referentie paper	Titel	Auteur	Jaar van publicatie	Aantal referenties in Web of Science	Methodologie paper, individueel experiment, niet van toepassing (M;IE;nvt)	Doel: link met intuïtieveit? (J;N)	Typen vragen: recall, comprehension, problem-solving, cloze (R;CO;PS;CL)	Soort respondenten: students vs business experts (ST;BE)
1	Comprehension of Procedural Visual Business Process Models A Literature Review	Figl, Kathrin	2017	20	M	J	R;CO;PS;CL	nvt
2	Exploring cognitive style and task-specific preferences for process representations	Figl, Kathrin; Recker, Jan	2016	18	IE	J	CO;PS	ST
3	Enhancing understandability of process models through cultural-dependent color adjustments	Kummer, Tyge-F.; Recker, Jan; Mendling, Jan	2016	15	nvt			
4	Lisbon Symbol Database (LSD): Subjective norms for 600 symbols	Prada, Marilia; Rodrigues, David; Silva, Rita R.; et al.	2016	12	nvt			
5	Learning from Quality Issues of BPMN Models from Industry	Leopold, Henrik; Mendling, Jan; Guenther, Oliver	2015	11	nvt			
6	Influence factors for local comprehensibility of process models	Figl, Kathrin; Laue, Ralf	2015	10	nvt			
7	How visual cognition influences process model comprehension	Petrusel, Razvan; Mendling, Jan; Reijers, Hajo A.	2017	9	nvt			
8	An empirical investigation of intuitive understandability of process diagrams	Jost, Gregor; Huber, Jemel; Hericko, Marjan; et al.	2016	8	IE	J	CO	ST
9	A Guidelines framework for understandable BPMN models	Corradini, Flavio; Ferrari, Alessio; Fomari, Fabrizio; et al.	2018	7	nvt			
10	Repeated use of process models: The impact of artifact, technological, and individual factors	Nolte, Alexander; Bernhard, Eike; Recker, Jan; et al.	2016	8	nvt			
11	The impact of working memory and the "process of process modelling" on model quality: Investigating experienced versus inexperienced modellers	Martini, Markus; Pinggera, Jakob; Neuraute, Manuel; et al.	2016	2	nvt			
12	The influence of directive explanations on users' business process compliance performance	Hadasch, Frank; Maedche, Alexander; Gregor, Shirley	2016	2	nvt			
13	An Empirical Review of the Connection Between Model Viewer Characteristics and the Comprehension of Conceptual Process Models	Mendling, Jan; Recker, Jan; Reijers, Hajo A.; et al.	2019	1	IE	N	CO	ST;BE
14	How do Individuals Interpret Multiple Conceptual Models? A Theory of Combined Ontological Completeness and Overlap	Recker, Jan; Green, Peter F.	2019	1	nvt			
15	Evaluation of Effective Cognition for the QGIS Processing Modeler	Dobesova, Zdena	2020	0	IE	N	CO	ST
16	USING EYE TRACKING TO EXPOSE COGNITIVE PROCESSES IN UNDERSTANDING CONCEPTUAL MODELS	Bera, Palash; Soffer, Phina; Parsons, Jeffrey	2019	0	IE	N	CO;PS	ST
17	Comprehension of business process models: insight into cognitive strategies via eye tracking	Tallon, Miles; Winter, Michael; Prys, s, Ruediger; et al.	2019	0	IE	N	CO	ST;BE
18	How quickly do we learn conceptual models?	Bera, Palash; Poels, Geert	2019	0	nvt			

Tabel 21: Analyseresultaten forward snowballing Moody et al. (2010)

Referentie paper	Titel	Auteur	Jaar van publicatie	Aantal referenties in Web of Science	Methodologie paper, individueel experiment, niet van toepassing (M;IE;nvt)	Doel: link met intuïtiviteit? (J;N)	Typen vragen: recall, comprehension, problem-solving, cloze (R;CO;PS;CL)	Soort respondenten: students vs business experts (ST;BE)
1	User requirements modeling and analysis of software-intensive systems	Soares, Michel dos Santos; Vrandeken, Jos; Verbraeck, Alexander	2011	20	nvt			
2	Visual requirements analytics: a framework and case study	Reddivari, Sandeep; Rad, Shirin; Bhowmik, Tanmay; et al.	2014	16	nvt			
3	The Use of Visualization in the Communication of Business Strategies: An Experimental Evaluation	Kernbach, Sebastian; Eppler, Martin J.; Bresciani, Sabrina	2015	13	nvt			
4	A scientific evaluation of the misuse case diagrams visual syntax	Saleh, Faisal; El-Attar, Mohamed	2015	12	IE	N	CO	BE
5	A questionnaire-based survey methodology for systematically validating goal-oriented models	Hassine, Jameleddine; Amyot, Daniel	2016	9	nvt			
6	An Empirical Evaluation of the Effectiveness of Attack Graphs and Fault Trees in Cyber-Attack Perception	Lallie, Harjinder Singh; Debatista, Kurt; Bal, Jay	2018	7	NA			
7	Empirical validating the cognitive effectiveness of a new feature diagrams visual syntax	Saeed, Mazin; Saleh, Faisal; Al-Insaf, Sadiq; et al.	2016	7	IE	N	CO	ST
8	Extracting conceptual models from user stories with Visual Narrator	Lucassen, Garm; Robeer, Marcel; Dalpiaz, Fabiano; et al.	2017	7	nvt			
9	Understanding what is important in iStar extension proposals: the viewpoint of researchers	Goncalves, Enyo; de Oliveira, Marcos Antonio; Monteiro, Ingrid; et al.	2019	2	nvt			
10	Improving cognitive effectiveness of business process diagrams with opacity-driven graphical highlights	Jost, Gregor; Huber, Jernej; Hericko, Marjan; et al.	2017	1	IE	N	CO	ST
11	A review of attack graph and attack tree visual syntax in cyber security	Lallie, Harjinder Singh; Debatista, Kurt; Bal, Jay	2020	0	nvt			
12	A comparative study of students and professionals in syntactical model comprehension experiments	El-Attar, Mohamed	2019	0	IE	N	CO	BE+ST
13	Semiotic clarity of Case Management Model and Notation (CMMN)	Bule, Mateja Kocbek; Polancic, Gregor; Huber, Jernej; et al.	2019	0	nvt			
14	Evaluating and empirically improving the visual syntax of use case diagrams	El-Attar, Mohamed	2019	0	IE	N	CO	BE
15	A Systematic Literature Review of Applications of the Physics of Notations	van der Linden, Dirk; Hadar, Irit	2019	0	nvt			
16	Theoretical foundations and implementation of business process diagrams' complexity management technique based on highlights	Jost, Gregor; Hericko, Marjan; Polancic, Gregor	2019	0	nvt			
17	A Semantic Approach to Deploying Product-Service Systems	Buchmann, Robert Andrei; Karagiannis, Dimitris	2017	0	nvt			

Tabel 22: Analyseresultaten forward snowballing Figl (2017)

Referentie paper	Titel	Auteur	Jaar van publicatie	Aantal referenties in Web of Science	Methodologie paper, individueel experiment, niet van toepassing (M;IE;nvt)	Doel: link met intuïtiviteit? (J;N)	Typen vragen: recall, comprehension, problem-solving, cloze (R;CO;PS;CL)	Soort respondenten: students vs business experts (ST;BE)
1	A Guidelines framework for understandable BPMN models	Corradini, Flavio; Ferrari, Alessio; Fornari, Fabrizio; et al.	2018	8	nvt			
2	On the declarative paradigm in hybrid business process representations: A conceptual framework and a systematic literature study	Andaloussi, Amine Abbad; Burattin, Andrea; Slaats, Tijs; et al.	2020	1	nvt			
3	The Influence of Using Collapsed Sub-processes and Groups on the Understandability of Business Process Models	Turetken, Oktay; Dikici, Ahmet; Vanderfeesten, Irene; et al.	2020	1	IE	N	CO	BE+ST
4	An Empirical Review of the Connection Between Model Viewer Characteristics and the Comprehension of Conceptual Process Models	Mendling, Jan; Recker, Jan; Reijers, Hajo A.; et al.	2019	1	nvt			
5	Investigating styles in variability modeling: Hierarchical vs. constrained styles	Reinhartz-Berger, Iris; Figl, Kathrin; Haugen, Oystein	2017	1	nvt			
6	An experimental design on the SPEM 2.0 process model element classification algorithm of the AVISPA tool through ANOVA variance analysis	Alvarez Londono, Jhon Jairo; Hurtado Alegria, Julio Ariet	2020	0	nvt			
7	A requirements engineering methodology for knowledge management solutions: integrating technical and social aspects	Levy, Meira; Hadar, Irit; Aviv, Itzhak	2019	0	nvt			
8	Comprehension of business process models: Insight into cognitive strategies via eye tracking	Tallon, Miles; Winter, Michael; Pryss, Ruediger; et al.	2019	0	nvt			
9	A novel approach to measuring enterprise procurement decision process: an information distance perspective	Li, Xiong; Zhao, Xiaodong; Pu, Wei	2020	0	nvt			
10	How product representation influences the understanding of supply chain process models	Leukel, Joerg; Sugumaran, Vijayan	2018	0	nvt			

Bijlage 3: Samenvatting papers theoretisch kader

Aranda et al. (2007)

Aranda et al. (2007) presenteren een theoretisch en methodologisch gegrond raamwerk voor het empirisch evalueren van de begrijpelijkheid (i.e. comprehensibility) van procesmodellen. Deze afhankelijke variabele wordt gemeten met zowel objectieve als subjectieve maatstaven. De objectieve maatstaven richten zich op de mate waarin participanten begripstaken correct hebben beantwoord, oftewel de interpretatieve effectiviteit (i.e. correctness of understanding). Daarnaast wordt met de interpretatieve efficiëntie (i.e. time) gemeten hoeveel tijd de participant nodig heeft om de begripstaken af te ronden. Voor de subjectieve maatstaven worden vragen beantwoord over in hoeverre de participant vertrouwen heeft in de juistheid van het gegeven antwoord bij begripstaken - ook wel de subjectieve interpretatieve effectiviteit (i.e. confidence) genoemd - en over het waargenomen gebruiksgemak (i.e. perceived difficulty) bij het beantwoorden van de begripvragen. De onafhankelijke variabelen zijn taaksoort (i.e. type of task), modelleerkennis (i.e. language expertise), domeinbegrip (i.e. domain expertise) en probleemomvang (i.e. problem size).

Figl (2017)

Figl (2017) geeft in het artikel inzicht in de bronnen van cognitieve inspanning bij de begrijpelijkheid (i.e. comprehension) van procesmodellen en stelt een reeks onafhankelijke variabelen voor die invloed hierop hebben. De afhankelijke variabele wordt gemeten door de objectieve maatstaven interpretatieve effectiviteit (i.e. comprehension accuracy) en de interpretatieve efficiëntie (i.e. time taken). De subjectieve maatstaven zijn de waargenomen gebruiksgemak (i.e. comprehension difficulty) en de mate waarin de gebruiker een voorkeur aangeeft voor een specifieke notatie boven een andere notatie, oftewel representatievoorkeur (i.e. user preference). De onafhankelijke variabelen zijn (i) presentatiemedium (i.e. presentation medium); (ii) notatie (i.e. notation); (iii) secundaire notatie (i.e. secondary notation); (iv) naamgeving (i.e. label); (v) modeleigenschappen (i.e. model characteristics); (vi) taaksoort (i.e. task); en (vii) gebruikerseigenschappen (i.e. user). De variabele (i) presentatiemedium veronderstelt een verschil in begrijpelijkheid bij de manier waarop de notatie wordt aangeboden (i.e. digitaal versus op papier). De (ii) notatie bestaat uit drie factoren: het representatie paradigma (i.e. representation paradigm), dat zich richt op verschillen in begrijpelijkheid tussen representatievormen (e.g. tekst versus model); de primaire notatie (i.e. primary notation), die zich richt op verschillen in begrijpelijkheid tussen modelleertalen als geheel (e.g. BPMN versus UML); en de notatie-eigenschappen (i.e. notational characteristics), die zich richten op verschillen in begrijpelijkheid tussen specifieke notatie-eigenschappen (e.g. BPMN gateways versus UML gateways). Daar waar de primaire notatie gaat over de formele definities van symbolen en regels over hoe deze gecombineerd kunnen worden, gaat de (iii) secundaire notatie over de visuele variabelen die niet formeel gespecificeerd zijn. De secundaire notatie bestaat uit de factoren decompositieregels (i.e. decomposition); Gestalt theorie; en layout. Bij decompositieregels wordt een verschil in begrijpelijkheid verwacht bij de mate waarin een notatie werkt met sub-modellen, waardoor de gebruiker bij grote modellen niet wordt blootgesteld aan onnodige informatie. De Gestalt theorie gaat ervan uit dat begrijpelijkheid toeneemt wanneer de gebruiker eenvoudig kan zien dat aan elkaar gerelateerde elementen bij elkaar horen (e.g. door elementen te clusteren met kleuren of swim lanes). Layout heeft van doen met de mate waarin begrijpelijkheid wordt beïnvloed door layout eigenschappen, zoals modelleerrichting en symboolgrootte. Variabele (iv) naamgeving bestaat uit de factoren label design en naamgevingsconventies (i.e. naming conventions). Label design richt zich op verschillen in begrijpelijkheid tussen activiteiten met tekstuele en abstracte labels. Naamgevingsconventies kijken naar het verschil in begrijpelijkheid bij verschillende tekstuele zinsconstructies (e.g. woordvolgorde, aantal woorden). De (v) modeleigenschappen variabele bestaat

uit de factoren maat (i.e. size measures); structuur (i.e. structuredness), gateway-samenspel (i.e. gateway interplay); connecties (i.e. connection); en syntax regels (i.e. syntax rules). De maat richt zich op begrijpelijkheidsverschillen bij het aantal notatie-elementen. Structuur richt zich op de mate waarin structuurverschillen (e.g. gebruik van nesting) in modellen effect hebben op de begrijpelijkheid ervan. Gateway-samenspel richt zich op verschillen in begrijpelijkheid tussen opeenvolgende gateways. Connecties gaat over de invloed die het aantal routeringspaden heeft op de begrijpelijkheid van procesmodellen. Syntax regels veronderstelt een relatie tussen onder andere het aantal syntax regels en de begrijpelijkheid van de notatie. Bij de (vi) taaksoort wordt onderzocht wat de invloed is van het type taak, alsook andere taakeigenschappen op de begrijpelijkheid van de notatie. Variabele (vii) gebruikerseigenschappen wordt beschreven als de benodigde vaardigheden, kennis en ervaring die de gebruiker zou moeten hebben om de notatie te kunnen begrijpen. De variabele bestaat uit de factoren domeinbegrip (i.e. domain knowledge), modelleerervaring (i.e. experience and familiarity with modeling), modelleerkennis (i.e. modeling knowledge) en opleidingsniveau (i.e. education).

Moody et al. (2010)

Moody et al. (2010) onderzoeken de begrijpelijkheid (i.e. cognitive effectiveness) van een notatie (i.e. i* notatie) door met desk research een syntaxanalyse uit te voeren volgens de principes van Physics of Notations (Moody, 2009), waarmee cognitief effectieve notaties kunnen worden ontworpen. De afhankelijke variabele begrijpelijkheid wordt gemeten met de objectieve maatstaven interpretatieve efficiëntie (i.e. speed), interpretatieve effectiviteit (i.e. accuracy) en door de subjectieve maatstaf waargenomen gebruiksgemak (i.e. ease). Volgens Moody et al. (2010) zijn notaties begrijpelijk wanneer voldoende invulling wordt gegeven aan de in tabel 27 verder toegelichte onafhankelijke variabelen semiotic clarity; perceptual discriminability; semantic transparency; complexity management; cognitive integration; visual expressiveness; dual coding; graphic economy; en cognitive fit.

Tabel 23: Onafhankelijke variabelen Moody et al. (2010)

Onafhankelijke variabele	Toelichting
Semiotic clarity	Semantische constructen en grafische symbolen moeten 1-op-1 met elkaar overeenkomen
Perceptual discriminability	Symbolen moeten onderling duidelijk van elkaar onderscheiden kunnen worden
Semantic transparency	Er moet gebruik gemaakt worden van symbolen die hun betekenis suggereren
Complexity management	Het opnemen van expliciete regels om met complexiteit binnen de notatie om te gaan
Cognitive integration	Het voorzien in expliciete mechanismen voor het ondersteunen van informatie-integratie van verschillende diagrammen
Visual expressiveness	De mate waarin gebruik gemaakt wordt van het volledige assortiment en de mogelijkheden van visuele variabelen
Dual coding	Het complementair gebruiken van tekst aan notaties
Graphic economy	Er moet rekening gehouden worden met het aantal verschillende grafische symbolen om de notatie cognitief beheersbaar te houden
Cognitive fit	Het gebruik van verschillende notatiedialecten voor verschillende taken en/of verschillende gebruikers

Jošt et al. (2016)

Jošt et al. (2016) onderzoeken in hoeverre de notatie en diagramcomplexiteit van invloed zijn op de intuïtieve begrijpelijkheid (i.e. intuitive understandability) van procesdiagrammen, door met een experiment de intuïtieve te evalueren van een viertal processen in drie verschillende notaties: Event-driven Process Chains (i.e. EPC), UML Activity Diagrams en BPMN. De afhankelijke variabele intuïtieve begrijpelijkheid wordt beschreven als *“het gemak waarmee een notatie direct kan worden begrepen door gebruikers, zonder enige voorkennis of training”* (p. 91). De variabele kan worden gemeten door de objectieve maatstaven interpretatieve effectiviteit (i.e. score obtained, in %); interpretatieve efficiëntie (i.e. time needed, in seconden); en interpretatieve doeltreffendheid, de ratio van interpretatieve effectiviteit en interpretatieve efficiëntie. De onafhankelijke variabelen zijn primaire notatie (i.e. notation) en diagramcomplexiteit (i.e. diagram complexity). In het onderzoek wordt in een between-subjects experiment met een survey vragen gesteld aan studenten, zonder voorkennis van de onderzochte notaties.

Figl and Recker (2016)

De paper van Figl and Recker (2016) bevat een exploratief onderzoek over in hoeverre de afhankelijke variabele representatievoorkeur (i.e. process representation preference) van participanten wordt beïnvloed door de onafhankelijke variabelen taaksoort (i.e. application task); de cognitieve stijl (i.e. cognitive style); en de modelleerervaring van de participant. De afhankelijke variabele representatievoorkeur bevat vier mogelijke representatievormen: tekst; gestructureerde tekst; diagram; en iconen. De onafhankelijke variabele taaksoort bevat de vier mogelijkheden: begripstaak; communicatistaak; uitvoeringstaak; en verbetering taak. Bij de cognitieve stijl wordt vastgesteld met welke representatievorm de participant het beste informatie kan verwerken. Hierbij kan de voorkeur uitgaan naar visuele representaties – waarbij een onderscheid wordt gemaakt tussen object style (i.e.








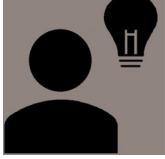

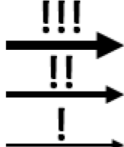
vorm en kleur) en spatial style (i.e. symbool locatie en ruimtelijke relaties); – en verbale representaties (i.e. woorden). Het onderzoek wordt uitgevoerd met een within-subjects experiment onder business school studenten, die representatief worden gezien als toekomstig eindgebruikers van de notaties. In een survey worden telkens twee verschillende notaties tegelijk getoond, waarbij de participant gevraagd wordt om met een slider de representatievoorkeur aan te geven.

Figl et al. (2013)

Figl et al. (2013) hebben onderzocht wat de invloed is van notatietekortkomingen (i.e. notational deficiencies) van een viertal notaties op de begrijpelijkheid (i.e. model comprehension) ervan. Dit is gedaan door de begrijpelijkheid van BPMN, EPC, UML Activity Diagrams en Yet Another Workflow Language (i.e. YAWL) met elkaar te vergelijken. De afhankelijke variabele begrijpelijkheid wordt gemeten met de objectieve maatstaven interpretatieve effectiviteit (i.e. model comprehension accuracy); interpretatieve efficiëntie (i.e. model comprehension efficiency); en de subjectieve maatstaf waargenomen gebruiksgemak (i.e. subjective cognitive load). De onafhankelijke variabelen zijn notatietekortkomingen op het gebied van semiotic clarity; en perceptual discriminability. In het onderzoek zijn tijdens een between-subjects experiment onder business school studenten, zonder voorkennis over procesmodelleren, begripstaken uitgevoerd.

Bijlage 4: 'Originele' en 'nieuwe' PGA-notatie elementen

Tabel 24: In dit onderzoek gebruikte 'originele' en 'nieuwe' PGA notaties

PGA-notatie element	'originele' versie	'nieuwe' versie
Customer goal		
Internal goal		
Value proposition		
Competence		
Value stream_importance		

Bijlage 5: Survey met experimentele taken (voor gebruik in Limesurvey)

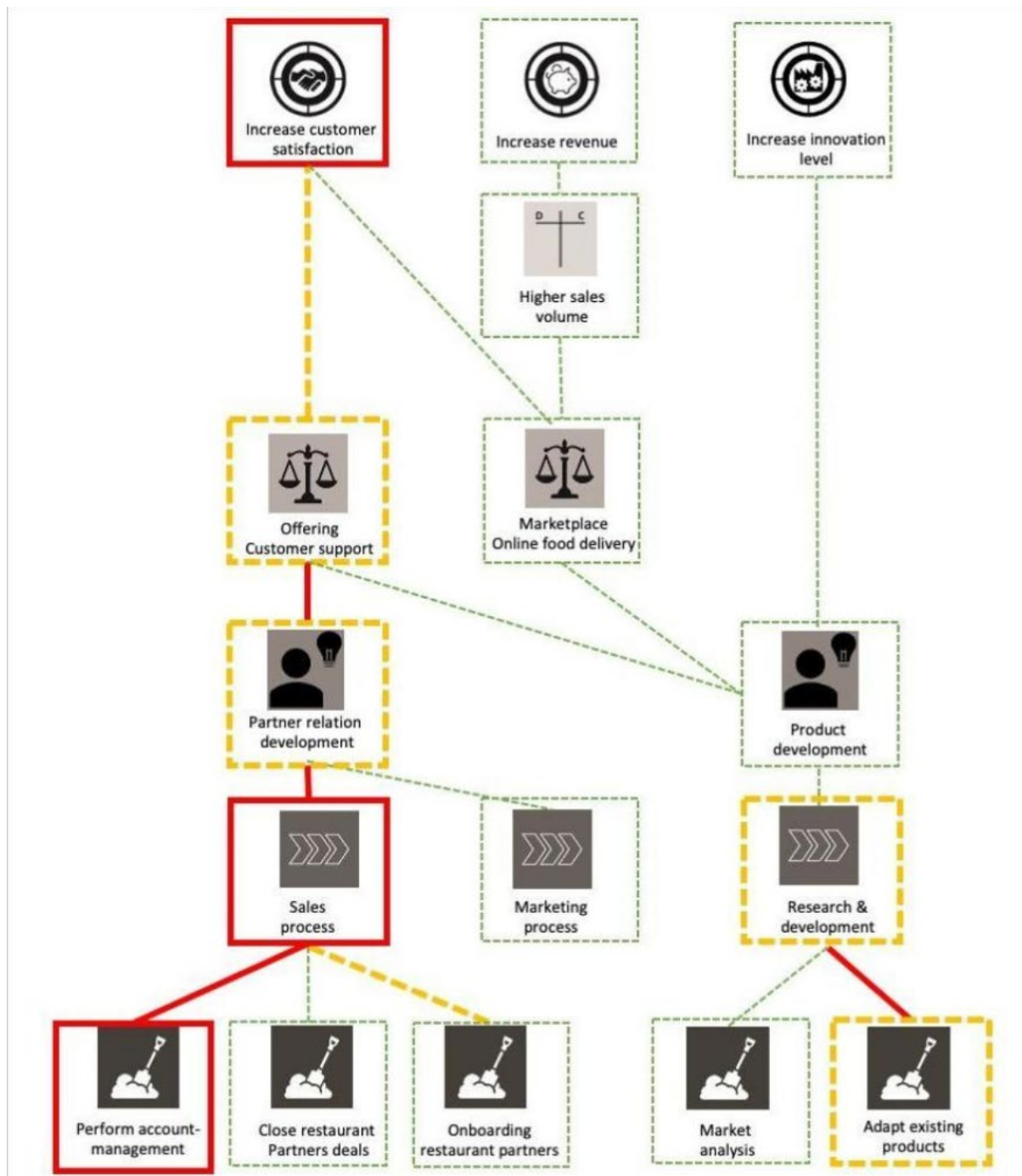


limesurvey_survey_5
47429.lss

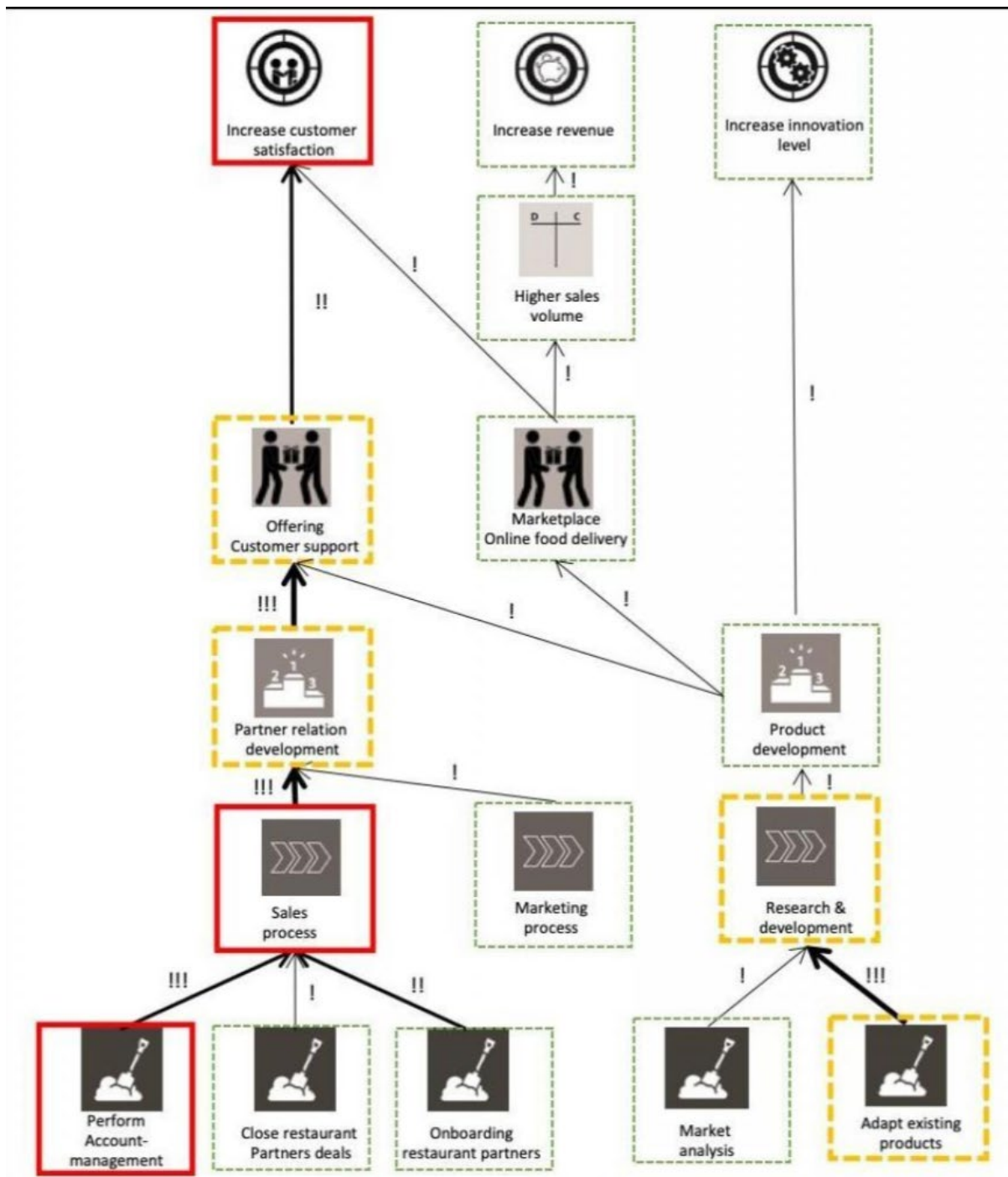
Bijlage 6: PGA-model voor begripstaken

Case description

The organization is a company that offers one leading marketplace for online food delivery and providing customers with an easy and secure way to order and pay for food from our restaurant partners.



Figuur 2: Casus voor survey versie A



Figuur 3: Casus voor survey versie B

Bijlage 7: Totaaloverzicht hypothesen

Tabel 25: Totaaloverzicht hypothesen interpretatieve effectiviteit

H_Effectiviteit	Hypothesen interpretatieve effectiviteit (i.e. 16)
1	De <u>interpretatieve effectiviteit</u> van de nieuwe PGA-notatie voor element internal goal tijdens de notatieassociatie taak is hoger dan die van de originele PGA-notatie voor element internal goal tijdens de notatieassociatie taak
2	De <u>interpretatieve effectiviteit</u> van de nieuwe PGA-notatie voor element value proposition tijdens de notatieassociatie taak is hoger dan die van de originele PGA-notatie voor element value proposition tijdens de notatieassociatie taak
3	De <u>interpretatieve effectiviteit</u> van de nieuwe PGA-notatie voor element competence tijdens de notatieassociatie taak is hoger dan die van de originele PGA-notatie voor element competence tijdens de notatieassociatie taak
4	De <u>interpretatieve effectiviteit</u> van de nieuwe PGA-notatie voor element value stream tijdens de notatieassociatie taak is hoger dan die van de originele PGA-notatie voor element value stream tijdens de notatieassociatie taak
5	De <u>interpretatieve effectiviteit</u> van de nieuwe PGA-notatie voor element customer goal tijdens de notatieassociatie taak is hoger dan die van de originele PGA-notatie voor element customer goal tijdens de notatieassociatie taak
6	De <u>interpretatieve effectiviteit</u> van de nieuwe PGA-notatie voor element customer goal tijdens de begripstaak is hoger dan die van de originele PGA-notatie voor element customer goal tijdens de begripstaak
7	De <u>interpretatieve effectiviteit</u> van de nieuwe PGA-notatie voor element value proposition tijdens de begripstaak is hoger dan die van de originele PGA-notatie voor element value proposition tijdens de begripstaak
8	De <u>interpretatieve effectiviteit</u> van de nieuwe PGA-notatie voor element value stream tijdens de begripstaak is hoger dan die van de originele PGA-notatie voor element value stream tijdens de begripstaak
9	De <u>interpretatieve effectiviteit</u> van de nieuwe PGA-notatie voor element competence tijdens de begripstaak is hoger dan die van de originele PGA-notatie voor element competence tijdens de begripstaak
10	De <u>interpretatieve effectiviteit</u> van de nieuwe PGA-notatie voor element internal goal tijdens de begripstaak is hoger dan die van de originele PGA-notatie voor element internal goal tijdens de begripstaak
11	De <u>interpretatieve effectiviteit</u> van de nieuwe PGA-notatie voor element importance tijdens de begripstaak is hoger dan die van de originele PGA-notatie voor element importance tijdens de begripstaak
12	De <u>interpretatieve effectiviteit</u> van de nieuwe PGA-notatie voor element internal goal tijdens de herinneringstaak is hoger dan die van de originele PGA-notatie voor element internal goal tijdens de herinneringstaak
13	De <u>interpretatieve effectiviteit</u> van de nieuwe PGA-notatie voor element value proposition tijdens de herinneringstaak is hoger dan die van de originele PGA-notatie voor element value proposition tijdens de herinneringstaak
14	De <u>interpretatieve effectiviteit</u> van de nieuwe PGA-notatie voor element competence tijdens de herinneringstaak is hoger dan die van de originele PGA-notatie voor element competence tijdens de herinneringstaak
15	De <u>interpretatieve effectiviteit</u> van de nieuwe PGA-notatie voor element value stream tijdens de herinneringstaak is hoger dan die van de originele PGA-notatie voor element value stream tijdens de herinneringstaak
16	De <u>interpretatieve effectiviteit</u> van de nieuwe PGA-notatie voor element customer goal tijdens de herinneringstaak is hoger dan die van de originele PGA-notatie voor element customer goal tijdens de herinneringstaak

Tabel 30: Totaaloverzicht hypothesen interpretatieve efficiëntie

H_Efficiëntie	Hypothesen interpretatieve efficiëntie (i.e. 16)
1	H_efficiëntie: de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element internal goal tijdens de notatieassociatie taak is hoger dan die van de originele PGA-notatie voor element internal goal tijdens de notatieassociatie taak
2	H_efficiëntie: de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element value proposition tijdens de notatieassociatie taak is hoger dan die van de originele PGA-notatie voor element value proposition tijdens de notatieassociatie taak
3	H_efficiëntie: de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element competence tijdens de notatieassociatie taak is hoger dan die van de originele PGA-notatie voor element competence tijdens de notatieassociatie taak
4	H_efficiëntie: de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element value stream tijdens de notatieassociatie taak is hoger dan die van de originele PGA-notatie voor element value stream tijdens de notatieassociatie taak
5	H_efficiëntie: de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element customer goal tijdens de notatieassociatie taak is hoger dan die van de originele PGA-notatie voor element customer goal tijdens de notatieassociatie taak
6	H_efficiëntie: de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element customer goal tijdens de begripstaak is hoger dan die van de originele PGA-notatie voor element customer goal tijdens de begripstaak
7	H_efficiëntie: de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element value proposition tijdens de begripstaak is hoger dan die van de originele PGA-notatie voor element value proposition tijdens de begripstaak
8	H_efficiëntie: de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element value stream tijdens de begripstaak is hoger dan die van de originele PGA-notatie voor element value stream tijdens de begripstaak
9	H_efficiëntie: de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element competence tijdens de begripstaak is hoger dan die van de originele PGA-notatie voor element competence tijdens de begripstaak
10	H_efficiëntie: de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element internal goal tijdens de begripstaak is hoger dan die van de originele PGA-notatie voor element internal goal tijdens de begripstaak
11	H_efficiëntie: de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element importance tijdens de begripstaak is hoger dan die van de originele PGA-notatie voor element importance tijdens de begripstaak
12	H_efficiëntie: de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element internal goal tijdens de herinneringstaak is hoger dan die van de originele PGA-notatie voor element internal goal tijdens de herinneringstaak
13	H_efficiëntie: de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element value proposition tijdens de herinneringstaak is hoger dan die van de originele PGA-notatie voor element value proposition tijdens de herinneringstaak
14	H_efficiëntie: de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element competence tijdens de herinneringstaak is hoger dan die van de originele PGA-notatie voor element competence tijdens de herinneringstaak
15	H_efficiëntie: de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element value stream tijdens de herinneringstaak is hoger dan die van de originele PGA-notatie voor element value stream tijdens de herinneringstaak
16	H_efficiëntie: de <u>interpretatieve efficiëntie</u> van nieuwe PGA-notatie voor element customer goal tijdens de herinneringstaak is hoger dan die van de originele PGA-notatie voor element customer goal tijdens de herinneringstaak

Tabel 31: Totaaloverzicht hypothesen representatievoorkeur

H_Representatievoorkeur	Hypothesen Representatievoorkeur (i.e. 5)
1	H_representatievoorkeur: de <u>representatievoorkeur</u> van de nieuwe PGA-notatie voor element customer goal tijdens de representatievoorkeur taak is hoger dan die van de originele PGA-notatie voor element customer goal tijdens de representatievoorkeur taak
2	H_representatievoorkeur: de <u>representatievoorkeur</u> van de nieuwe PGA-notatie voor element value proposition tijdens de representatievoorkeur taak is hoger dan die van de originele PGA-notatie voor element value proposition tijdens de representatievoorkeur taak
3	H_representatievoorkeur: de <u>representatievoorkeur</u> van de nieuwe PGA-notatie voor element competence tijdens de representatievoorkeur taak is hoger dan die van de originele PGA-notatie voor element competence tijdens de representatievoorkeur taak
4	H_representatievoorkeur: de <u>representatievoorkeur</u> van de nieuwe PGA-notatie voor element internal goal tijdens de representatievoorkeur taak is hoger dan die van de originele PGA-notatie voor element internal goal tijdens de representatievoorkeur taak
5	H_representatievoorkeur: de <u>representatievoorkeur</u> van de nieuwe PGA-notatie voor element value stream tijdens de representatievoorkeur taak is hoger dan die van de originele PGA-notatie voor element value stream tijdens de representatievoorkeur taak

Tabel 32: : Totaaloverzicht hypothesen construct intuïtieve begrijpelijkheid (i.e. geaggregeerde score van interpretatieve effectiviteit, -efficiëntie en representatievoorkeur)

H_Intuïtiviteit	Hypothese intuïtieve begrijpelijkheid (i.e. 5)
1	De <u>intuïtieve begrijpelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, -efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element internal goal is hoger dan die van de originele PGA-notatie voor element internal goal
2	De <u>intuïtieve begrijpelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, -efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element value proposition is hoger dan die van de originele PGA-notatie voor element value proposition
3	De <u>intuïtieve begrijpelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, -efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element competence is hoger dan die van de originele PGA-notatie voor element competence
4	De <u>intuïtieve begrijpelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, -efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element value stream is hoger dan die van de originele PGA-notatie voor element value stream
5	De <u>intuïtieve begrijpelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, -efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element customer goal is hoger dan die van de originele PGA-notatie voor element customer goal

Bijlage 8: Gedetailleerde analyse onderzoeksresultaten per element op taakniveau van de eigen dataset

Excel



SPSS_Score per
notatie op taakniveau

Bijgevoegd excel-bestand bevat een reeks tabbladen met de analyse van de onderzoeksresultaten per notatie op taakniveau (i.e. notatieassociatie taken, begripstaken, herinneringstaken en representatievoorkeur taken). Hiermee kan vervolgens op taakniveau m.b.v. hypothesetoetsen worden vastgesteld in hoeverre de 'nieuwe' notatie hoger scoort op interpretatieve effectiviteit, interpretatieve efficiëntie en representatievoorkeur – en dus als intuïtiever wordt ervaren in vergelijking met de 'originele'.

Voordat de oorspronkelijke data uit *Limesurvey* door SPSS kan worden verwerkt is de data allereerst gecodeerd, zie tabblad 'Codes' voor de gebruikte codering en 'SPSS_gecodeerde data_Merijn' voor de gecodeerde data. Voor de meeste variabelen (i.e. 'Version', 'Gender', 'Degree', 'JobSector', 'ModelingExperience', 'ModelingExpertise', 'ModelNumberRead', 'NumberModelsMade', 'Begripsvragen') geldt dat de schaal start bij 0 en vervolgens positieve waardes aanneemt. Alleen bij de notatieassociatie taken en herinneringstaken is ook een negatieve waarde mogelijk (i.e. -1), om extra gewicht toe te kennen aan het feit dat de respondent een 'ander meta model element' heeft genoemd. Hiermee wordt namelijk aangegeven dat de notaties onderling niet duidelijk van elkaar kunnen worden onderscheiden en dus dat de perceptual discriminability in het geding komt. Een andere bijzonderheid aan het coderen van notatieassociatie- en herinneringstaken is dat hierin een zekere mate van subjectiviteit van de onderzoeker komt kijken. De participant heeft bij deze onderdelen zelf namelijk tekst ingevoerd, waardoor door de onderzoeker zelf in moet schatten in welke mate de antwoorden overeenkomen met het antwoordmodel (i.e. 'ander meta model element', 'fout', 'gedeeltelijk' of 'goed').

'Resultaten_Merijn' bevat de hypothesen, een toelichting op de gekozen hypothesetoetsen en geeft ook de resultaten hiervan weer. Hierna volgen de tabbladen met toetsen die hieraan ten grondslag liggen: in 'Norm._schaal_Merijn' staan de resultaten van de normaliteitstoets op de schaalvariabelen, in 'Mann-Whitney Test_schaal_Merijn' de resultaten van de hypothesetests op de ongepaarde niet-normaal verdeelde schaalvariabelen en in 'Mann-Whitney Test_ordina_Merijn' die van de ongepaarde ordinale variabelen. In 'Ind.Sample T_schaal_Merijn' staan de resultaten van de ongepaarde normaal verdeelde schaalvariabelen. 'One Sample Wilcoxon_SR_Merijn' bevat de hypothesetests op de gepaarde niet-normaal verdeelde schaalvariabelen, in 'One Sample T_Merijn' die van de normaal verdeelde.

- 'nieuwe' notatie scoort beter dan 'originele' notatie
 - o (originele '2-tailed' significantie)/2
- 'originele' notatie scoort beter dan 'nieuwe' notatie
 - o (1-originele '2-tailed' significantie)/2

Zoals zichtbaar in tabel 33 is bij drie variabelen een significante afwijking te zien tussen de ‘nieuwe’ en ‘originele’ notaties. Zo is te zien dat de ‘nieuwe’ notatie voor value proposition en competence in de representatievoorkeur taak significant beter scoren dan de ‘originele’. Ook de is ‘nieuwe’ notatie voor internal goal bij de herinnertaak significant sneller geraden dan de ‘originele’.

Nulhypothese	Statistische test	Asymp. Sig. (1-tailed)	Significant
H <u>representatievoorkeur</u> : de representatievoorkeur van de nieuwe PGA-notatie voor element value proposition tijdens de representatievoorkeur taak is niet hoger dan die van de originele PGA-notatie voor element value proposition tijdens de representatievoorkeur taak	Wilcoxon Test, Sign Test	0.02	ja
H <u>representatievoorkeur</u> : de representatievoorkeur van de nieuwe PGA-notatie voor element competence tijdens de representatievoorkeur taak is niet hoger dan die van de originele PGA-notatie voor element competence tijdens de representatievoorkeur taak	One-Sample T-test	0.02	ja
H <u>efficiëntie</u> : de interpretatieve efficiëntie van nieuwe PGA-notatie voor element internal goal tijdens de herinneringstaak is niet hoger dan die van de originele PGA-notatie voor element internal goal tijdens de herinneringstaak	Wilcoxon-Mann-Whitney Test	0.01	ja

Naast de excel zijn ook de bestanden toegevoegd met input- en outputdata van SPSS:



68

Bijlage 9: Gedetailleerde analyse onderzoeksresultaten per element op taakniveau van de totale dataset

Excel



SPSS_Totaalscore per notatie op taakniveau.

Bijgevoegd excel-bestand bevat een reeks tabbladen met de analyse van de onderzoeksresultaten per notatie op taakniveau (i.e. notatieassociatie taken, begripstaken, herinneringstaken en representatievoorkeur taken). Hiermee kan vervolgens op taakniveau m.b.v. hypothesetoetsen worden vastgesteld in hoeverre de 'nieuwe' notatie hoger scoort op interpretatieve effectiviteit, interpretatieve efficiëntie en representatievoorkeur – en dus als intuïtiever wordt ervaren in vergelijking met de 'originele'.

Voordat de oorspronkelijke data uit *Limesurvey* door SPSS kan worden verwerkt is de data allereerst gecodeerd, zie tabblad 'Codes' voor de gebruikte codering en 'SPSS_gecodeerde data' voor de gecodeerde data. Voor de meeste variabelen (i.e. 'Version', 'Gender', 'Degree', 'JobSector', 'ModelingExperience', 'ModelingExpertise', 'ModelNumberRead', 'NumberModelsMade', 'Begripsvragen') geldt dat de schaal start bij 0 en vervolgens positieve waardes aanneemt. Alleen bij de notatieassociatie taken en herinneringstaken is ook een negatieve waarde mogelijk (i.e. -1), om extra gewicht toe te kennen aan het feit dat de respondent een 'ander meta model element' heeft genoemd. Hiermee wordt namelijk aangegeven dat de notaties onderling niet duidelijk van elkaar kunnen worden onderscheiden en dus dat de perceptual discriminability in het geding komt. Een andere bijzonderheid aan het coderen van notatieassociatie- en herinneringstaken is dat hierin een zekere mate van subjectiviteit van de onderzoeker komt kijken. De participant heeft bij deze onderdelen zelf namelijk tekst ingevoerd, waardoor door de onderzoeker zelf in moet schatten in welke mate de antwoorden overeenkomen met het antwoordmodel (i.e. 'ander meta model element', 'fout', 'gedeeltelijk' of 'goed').

'Resultaten_totaal' bevat de hypothesen, een toelichting op de gekozen hypothesetoetsen en geeft ook de resultaten hiervan weer. Hierna volgen de tabbladen met toetsen die hieraan ten grondslag liggen: in 'Norm._schaal_totaal' staan de resultaten van de normaliteitstoets op de schaalvariabelen, in 'Mann-Whitney Test_schaal_totaal' de resultaten van de hypothesetesten op de ongepaarde niet-normaal verdeelde schaalvariabelen en in 'Mann-Whitney Test_ordina_totaal' die van de ongepaarde ordinale variabelen. 'One Sample Wilcoxon_SR_totaal' bevat de hypothesetesten op de gepaarde niet-normaal verdeelde schaalvariabelen, in 'One Sample T_totaal' die van de normaal verdeelde.

Aangezien het onderzoek zich richt op de mate waarin de 'nieuwe' notatie als intuïtiever wordt beoordeeld dan de 'originele' moet bij de hypothesetoetsen een aanvullende berekening worden uitgevoerd. Normaal gezien kan gekeken worden naar de '2-tailed' significantie, omdat de richting van de significantie niet van belang is. In geval van dit onderzoek is die wel belangrijk: onderzocht wordt namelijk of de 'nieuwe' notatie significant beter scoort dan de 'originele'. Andersom is niet van belang, vandaar dat de volgende berekening moet worden uitgevoerd om de '2-tailed' significantie om te zetten naar de '1-tailed' significantie:

- 'nieuwe' notatie scoort beter dan 'originele' notatie
 - o (originele '2-tailed' significantie)/2
- 'originele' notatie scoort beter dan 'nieuwe' notatie
 - o (1-originele '2-tailed' significantie)/2

Resultaten

Zoals zichtbaar in tabel 34 is bij zes variabelen een significante afwijking te zien tussen de 'nieuwe' en 'originele' notaties. Zo is te zien dat de 'nieuwe' notatie voor 'Value proposition' in twee taken (i.e. notatieassociatie en representatievoorkeur) en 'Customer goal' in één taak (i.e. representatievoorkeur) significant beter scoort dan de 'originele'. Ook de 'nieuwe' notaties voor 'internal goal' en 'competence' zijn bij de herinnertaak significant sneller geraden dan de 'originele'. Tot slot is de 'nieuwe' notatie voor 'importance' significant sneller begrepen in de begripstaak dan de 'originele'.

Tabel 34: Resultaten intuïtieve begripelijkheid per notatie op taakniveau (volledige dataset)

Nulhypothese	Statistische test <i>boven: eigen set; onder: totale set</i>	Asymp. Sig. (1-tailed) <i>boven: eigen set; onder: totale set</i>	Significant <i>boven: eigen set; onder: totale set</i>
H representatievoorkeur: de representatievoorkeur van de nieuwe PGA-notatie voor element value proposition tijdens de representatievoorkeur taak is niet hoger dan die van de originele PGA-notatie voor element value proposition tijdens de representatievoorkeur taak	Wilcoxon Test, Sign Test	0.0005	ja
H representatievoorkeur: de representatievoorkeur van de nieuwe PGA-notatie voor element customer goal tijdens de representatievoorkeur taak is niet hoger dan die van de originele PGA-notatie voor element customer goal tijdens de representatievoorkeur taak	Wilcoxon Test, Sign Test	0.021	ja
H efficiëntie: de interpretatieve efficiëntie van nieuwe PGA-notatie voor element internal goal tijdens de herinneringstaak is niet hoger dan die van de originele PGA-notatie voor element internal goal tijdens de herinneringstaak	Wilcoxon-Mann-Whitney Test	0.045	ja
H efficiëntie: de interpretatieve efficiëntie van nieuwe PGA-notatie voor element competence tijdens de herinneringstaak is niet hoger dan die van de originele PGA-notatie voor element competence tijdens de herinneringstaak	Wilcoxon-Mann-Whitney Test	0.0495	ja
H efficiëntie: de interpretatieve efficiëntie van nieuwe PGA-notatie voor element importance tijdens de begripstaak is niet hoger dan die van de originele PGA-notatie voor element importance tijdens de begripstaak	Wilcoxon-Mann-Whitney Test	0.004	ja
H effectiviteit: de interpretatieve effectiviteit van de nieuwe PGA-notatie voor element value proposition tijdens de notatieassociatie taak is niet hoger dan die van de originele PGA-notatie voor element value proposition tijdens de notatieassociatie taak	Wilcoxon-Mann-Whitney Test	0.026	ja

SPSS

Naast de excel zijn ook de bestanden toegevoegd met input- en outputdata van SPSS:



SPSS_testoutput_Tot
aal.spv



SPSS_input_data_Tot
aal.sav

Bijlage 10: Gedetailleerde analyse geaggregeerde onderzoeksresultaten intuïtiviteit van de eigen dataset

Excel



SPSS_Geaggregeerde
totaalscore per notatie

Bijgevoegd excel-bestand bevat een reeks tabbladen met de analyse van de onderzoeksresultaten van de eigen dataset voor het hele construct intuïtiviteit, waarbij de resultaatscores van de verschillende taken (i.e. notatieassociatie taken, begripstaken, herinneringstaken en representatievoorkeur taken) bij elkaar worden opgeteld, om zodoende tot één score per participant, per notatie te komen. Hiermee kan vervolgens m.b.v. hypothesetoetsen worden vastgesteld in hoeverre de 'nieuwe' notatie hoger scoort op intuïtieve begrijpelijkheid in vergelijking met de 'originele'. als intuïtiever wordt ervaren in vergelijking met de 'originele'.

Echter, de taakscores kunnen niet zomaar bij elkaar worden opgeteld, omdat deze verschillende schalen bevatten. Daarom moeten de originele scores herschaald worden naar één schaal, namelijk naar percentages (zie tabblad 'Taakscores herschaald in % .Me'). Voorafgaand hieraan moeten enkele variabelen (i.e. notatieassociatie taken en herinneringstaken) opnieuw gecodeerd worden (zie tabblad 'Codes_Geaggr.'), zodat deze geen negatieve waarden meer kunnen bevatten.

Om tot de nieuwe codering en schaling te komen zijn de volgende berekeningen gebruikt:

- Notatieassociatie taken
 - o $((\text{originele score} + 1) / 3) * 100\%$
 - Ivm opnieuw coderen
 - originele score +1
 - Ivm herschalen naar percentage delen door maximumscore
 - voorgaande score / 3
- Begripstaken
 - o $(\text{originele score}) * 100\%$
 - Originele score kan 0 of 1 zijn
 - Ivm herschalen naar percentage * 100%
- Herinneringstaken
 - o $((\text{originele score} + 1) / 3) * 100\%$
 - Ivm opnieuw coderen
 - originele score +1
 - Ivm herschalen naar percentage delen door maximumscore
 - voorgaande score / 3
- Representatievoorkeur taken
 - o herschaling naar % tov versie voorgaande taken
 - Versie voorgaande taken = 'old'
 - $(100 - \text{originele score}) * 100\%$
 - Versie voorgaande taken = 'new'
 - $(\text{originele score}) * 100\%$

Vervolgens zijn in tabblad 'Geaggr.tot.score per notatie.Me' de scores uit voorgaand tabblad per participant, per notatie bij elkaar opgeteld. Voor de vier notaties internal goal, value proposition, competence en customer goal geldt dat de vier taakscores direct bij elkaar zijn opgeteld. In het geval van value stream is voor de begripstaken het gemiddelde is genomen van de value stream en importance notaties. Het betreft immers dezelfde notatie en wordt in andere taken ook niet afzonderlijk uitgevraagd. 'Resultaten_geaggr.tot.Merijn' bevat de hypothesen, een toelichting op de gekozen hypothesetoetsen en geeft ook de resultaten hiervan weer. Hierna volgen de tabbladen met toetsen die hieraan ten grondslag liggen: in 'Norm._geaggr.Merijn' staan de resultaten van de normaliteitstoets, in 'Mann-Whitney Test_geaggr.Merijn' de resultaten van de hypothesetesten op de niet-normaal verdeelde variabelen en in 'Ind.Sample T_geaggr.Merijn' die van de normaal verdeelde variabelen.

Aangezien het onderzoek zich richt op de mate waarin de 'nieuwe' notatie als intuïtiever wordt beoordeeld dan de 'originele' moet bij de hypothesetoetsen een aanvullende berekening worden uitgevoerd. Normaal gezien kan gekeken worden naar de '2-tailed' significantie, omdat de richting van de significantie niet van belang is. In geval van dit onderzoek is die wel belangrijk: onderzocht wordt namelijk of de 'nieuwe' notatie significant beter scoort dan de 'originele'. Andersom is niet van belang, vandaar dat de volgende berekening moet worden uitgevoerd om de '2-tailed' significantie om te zetten naar de '1-tailed' significantie:

- 'nieuwe' notatie scoort beter dan 'originele' notatie
 - o $(\text{originele '2-tailed' significantie})/2$
- 'originele' notatie scoort beter dan 'nieuwe' notatie
 - o $(1 - \text{originele '2-tailed' significantie})/2$

Resultaten



Zoals zichtbaar is in tabel 35 is geen significante afwijking te zien tussen de intuïtieve begrijpelijkheid van de 'nieuwe' en 'originele' notaties. Hiermee lijkt het erop dat participanten de 'nieuwe' notatie niet als intuïtiever beoordelen dan de 'originele' notatie.

Tabel 35: Geaggregeerde resultaten intuïtieve begrijpelijkheid per notatie (eigen dataset)

Nulhypothese	Statistische test	Asymp. Sig. (1-tailed)	Significant
H_intuïtiviteit: de <u>intuïtieve begrijpelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, - efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element internal goal is niet hoger dan die van de originele PGA-notatie voor element internal goal	Wilcoxon-Mann-Whitney Test	0.41	nee
H_intuïtiviteit: de <u>intuïtieve begrijpelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, - efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element value proposition is niet hoger dan die van de originele PGA-notatie voor element value proposition	Independent-samples T test	0.1525	nee
H_intuïtiviteit: de <u>intuïtieve begrijpelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, - efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element competence is niet hoger dan die van de originele PGA-notatie voor element competence	Independent-samples T test	0.407	nee
H_intuïtiviteit: de <u>intuïtieve begrijpelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, - efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element value stream is niet hoger dan die van de originele PGA-notatie voor element value stream	Independent-samples T test	0.4755	nee
H_intuïtiviteit: de <u>intuïtieve begrijpelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, - efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element customer goal is niet hoger dan die van de originele PGA-notatie voor element customer goal	Wilcoxon-Mann-Whitney Test	0.35	nee

SPSS

Naast de excel zijn ook de bestanden toegevoegd met input- en outputdata van SPSS:

 
 SPSS_testoutput_gea SPSS_input_data_gea
 ggr_Merijn.spv ggr_Merijn.sav

Bijlage 11: Gedetailleerde analyse geaggregeerde onderzoeksresultaten intuïtiviteit van de totale dataset

Excel



SPSS_Geaggregeerde
totaalscore per notatie

Bijgevoegd excel-bestand bevat een reeks tabbladen met de analyse van de onderzoeksresultaten van de totale dataset voor het hele construct intuïtiviteit, waarbij de resultaatscores van de verschillende taken (i.e. notatieassociatie taken, begripstaken, herinneringstaken en representatievoorkeur taken) bij elkaar worden opgeteld, om zodoende tot één score per participant, per notatie te komen. Hiermee kan vervolgens m.b.v. hypothesetoetsen worden vastgesteld in hoeverre de 'nieuwe' notatie hoger scoort op intuïtieve begrijpelijkheid in vergelijking met de 'originele'.

Echter, de taakscores kunnen niet zomaar bij elkaar worden opgeteld, omdat deze verschillende schalen bevatten. Daarom moeten de originele scores herschaald worden naar één schaal, namelijk naar percentages (zie tabblad bevat 'Taakscores herschaald in %'). Voorafgaand hieraan moeten enkele variabelen (i.e. notatieassociatie taken en herinneringstaken) opnieuw gecodeerd worden (zie tabblad 'Codes_Geaggr.'), zodat deze geen negatieve waarden meer kunnen bevatten.

Om tot de nieuwe codering en schaling te komen zijn de volgende berekeningen gebruikt:

- Notatieassociatie taken
 - o $((\text{originele score} + 1) / 3) * 100\%$
 - lvm opnieuw coderen
 - originele score +1
 - lvm herschalen naar percentage delen door maximumscore
 - voorgaande score / 3
- Begripstaken
 - o $(\text{originele score}) * 100\%$
 - Originele score kan 0 of 1 zijn
 - lvm herschalen naar percentage * 100%
- Herinneringstaken
 - o $((\text{originele score} + 1) / 3) * 100\%$
 - lvm opnieuw coderen
 - originele score +1
 - lvm herschalen naar percentage delen door maximumscore
 - voorgaande score / 3
- Representatievoorkeur taken
 - o herschaling naar % tov versie voorgaande taken
 - Versie voorgaande taken = 'old'
 - $(100 - \text{originele score}) * 100\%$
 - Versie voorgaande taken = 'new'
 - $(\text{originele score}) * 100\%$

Vervolgens zijn in tabblad 'Geaggr.tot.score per notatie' de scores uit voorgaand tabblad per participant, per notatie bij elkaar opgeteld. Voor de vier notaties 'Internal goal', 'Value proposition',

‘Competence’ en ‘Customer goal’ geldt dat de vier taakscores direct bij elkaar zijn opgeteld. In het geval van ‘Value stream’ is voor de begripstaken het gemiddelde is genomen van de ‘Value stream’ en ‘Importance’ notaties. Het betreft immers dezelfde notatie en wordt in andere taken ook niet afzonderlijk uitgevraagd. ‘Resultaten_geaggr.tot.’ bevat een toelichting op de gekozen hypothesetoetsen en geeft ook de resultaten hiervan weer. Hierna volgen de tabbladen met toetsen die hieraan ten grondslag liggen: in ‘Norm._geaggr.tot.’ staan de resultaten van de normaliteitstoets, in ‘Mann-Whitney Test_geaggr.tot.’ de resultaten van de hypothesetesten op de niet-normaal verdeelde variabelen en in ‘Ind.Sample T_geaggr.tot.’ die van de normaal verdeelde variabelen.

Aangezien het onderzoek zich richt op de mate waarin de ‘nieuwe’ notatie als intuïtiever wordt beoordeeld dan de ‘originele’ moet bij de hypothesetoetsen een aanvullende berekening worden uitgevoerd. Normaal gezien kan gekeken worden naar de ‘2-tailed’ significantie, omdat de richting van de significantie niet van belang is. In geval van dit onderzoek is die wel belangrijk: onderzocht wordt namelijk of de ‘nieuwe’ notatie significant beter scoort dan de ‘originele’. Andersom is niet van belang, vandaar dat de volgende berekening moet worden uitgevoerd om de ‘2-tailed’ significantie om te zetten naar de ‘1-tailed’ significantie:

- ‘nieuwe’ notatie scoort beter dan ‘originele’ notatie
 - o $(\text{originele '2-tailed' significantie})/2$
- ‘originele’ notatie scoort beter dan ‘nieuwe’ notatie
 - o $(1 - \text{originele '2-tailed' significantie})/2$

Resultaten

Zoals zichtbaar is in tabel 36 is geen significante afwijking te zien tussen de intuïtieve begrijpelijkheid van de 'nieuwe' en 'originele' notaties. Hiermee lijkt het erop dat participanten de 'nieuwe' notatie niet als intuïtiever beoordelen dan de 'originele' notatie.

Tabel 36: Geaggregeerde resultaten intuïtieve begrijpelijkheid per notatie (totale dataset)

Nulhypothese	Statistische test	Asymp. Sig. (1-tailed)	Significant
H_intuïtiviteit: de <u>intuïtieve begrijpelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, - efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element internal goal is niet hoger dan die van de originele PGA-notatie voor element internal goal	Wilcoxon-Mann-Whitney Test	0.16	nee
H_intuïtiviteit: de <u>intuïtieve begrijpelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, - efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element value proposition is niet hoger dan die van de originele PGA-notatie voor element value proposition	Independent-samples T test	0.3165	nee
H_intuïtiviteit: de <u>intuïtieve begrijpelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, - efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element competence is niet hoger dan die van de originele PGA-notatie voor element competence	Independent-samples T test	0.197	nee
H_intuïtiviteit: de <u>intuïtieve begrijpelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, - efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element value stream is niet hoger dan die van de originele PGA-notatie voor element value stream	Wilcoxon-Mann-Whitney Test	0.40	nee
H_intuïtiviteit: de <u>intuïtieve begrijpelijkheid</u> (i.e. geaggregeerde score van interpretatieve effectiviteit, - efficiëntie en representatievoorkeur) van de nieuwe PGA-notatie voor element customer goal is niet hoger dan die van de originele PGA-notatie voor element customer goal	Wilcoxon-Mann-Whitney Test	0.13	nee

SPSS

Naast de excel zijn ook de bestanden toegevoegd met input- en outputdata van SPSS:



SPSS_testoutput_gea
ggr_Totaal.spv



SPSS_input_data_gea
ggr_Totaal.sav

Bijlage 12: Gedetailleerde analyse kwalitatieve feedback representatievoorkeur

Excel



SPSS_Totaalscore per notatie op taakniveau

Bijgevoegd excel-bestand bevat een reeks tabbladen met per notatie de analyse van de kwalitatieve feedback (i.e. tekstuele toelichting bij de representatievoorkeur taak). Hiermee kan vervolgens doormiddel van triangulatie worden beoordeeld in hoeverre de kwalitatieve feedback de resultaten van de statistische testen over de representatievoorkeur hypothesen ondersteunen om zodoende de betrouwbaarheid van de onderzoeksresultaten te vergroten.

Door middel van een thematische analyse zijn thema's en patronen in de kwalitatieve data geïdentificeerd. In de tabbladen 'CompetenceQual', 'ValuePropositionQual', 'CustomerGoalQual', 'InternalGoalQual' en 'ValueStreamQual' zijn per element de tekstuele data met een gelijksoortige betekenis gecodeerd (i.e. een woord of korte zin) zodat thema's onderling met elkaar kunnen worden vergeleken. Aangezien gebruik is gemaakt van een deductieve onderzoekaankpak worden per element twee a priori codes (i.e. één voor de nieuwe en één voor de originele) gebruikt die zijn afgeleid van de gebruikte symbolen in de oude en nieuwe notaties. Hierbij wordt specifiek aandacht besteedt aan de mate waarin er sprake is van contrasterende meningen over de aangebrachte verbeteringen. In bovenstaande tabbladen is per element, per toelichting geturfd hoe vaak de thema's voorkomen. Om de kwalitatieve data als ondersteunend te zien voor de representatievoorkeur voor de 'nieuwe' notatie van een element, worden de volgende criteria toegepast: (i) minimaal 11 keer (i.e. 25% van de totale populatie) moet een specifiek thema worden genoemd in de kwalitatieve data; (ii) vervolgens moet dit thema minimaal twee-derde keer (i.e. 67%) gebruikt worden ten gunste van de nieuwe notatie. Per element moet minimaal één van de twee thema's aan het criterium voldoen om als ondersteunend te worden gezien voor de representatievoorkeur voor de 'nieuwe' notatie.

Resultaten

In tabblad 'TOTAALOVERZICHT_KWAL' - en eveneens in tabel 37 - zijn de resultaten zichtbaar, die tot stand gekomen zijn op basis van de analyse, met inachtneming van de twee criteria. De tabel bevat diverse kolommen: 'Element' bevat de relevante elementen; 'Thema' bevat de a priori codes die zijn afgeleid van de gebruikte symbolen in de oude en nieuwe notaties; '% thema genoemd t.o.v. totale populatie' bevat het percentage dat weergeeft of wordt voldaan aan criterium (i), inclusief het aantal keer dat het thema werd genoemd; '% contrasterende meningen t.v.v. nieuwe notatie' geeft het percentage weer dat iets zegt over de mate waarin wordt voldaan aan criterium (ii), inclusief informatie over de verhouding tussen het aantal keren dat het thema is gebruikt als argument om de nieuwe notatie te kiezen, versus het aantal keren dat voor de originele notatie is gekozen; 'Ondersteunend t.v.v. representatie-voorkeur nieuwe notatie' vertelt vervolgens per element, per thema wat de conclusie is, op basis van de gehanteerde criteria.

Immers werd bij element value proposition 18 keer (i.e. 41% van de totale populatie) het thema cadeau genoemd, waarbij 15 keer is aangegeven dat de 'nieuwe' notatie boven de 'originele' is verkozen omdat in de 'nieuwe' notatie het cadeau symbool met waarde werd geassocieerd, tegenover 3 contrasterende commentaren dat dit symbool juist niet met de betekenis van het element wordt geassocieerd. Dit komt erop neer dat 83% van de contrasterende meningen zijn gegeven ten voordele

van de 'nieuwe' notatie. Daarnaast komt thema schaal 14 keer (i.e. 32% van de totale populatie) in de data voor, alle 14 genoemd ten voordele van de 'nieuwe' notatie, omdat bij het schaalsymbool in de oude notatie geen link werd gezien met de betekenis van het element. Dit komt erop neer dat 100% van de contrasterende meningen zijn gegeven ten voordele van de 'nieuwe' notatie. Bij de customer goal is in totaal 20 keer (i.e. 45% van de totale populatie) het thema stakeholders genoemd. Hierbij werd 17 keer aangegeven dat de 'nieuwe' notatie als intuïtiever wordt bestempeld vanwege het feit dat er twee verschillende (i.e. business en non-business) stakeholders in te zien zijn waarin de klantrelatie wordt herkend, ten opzichte van 3 contrasterende meningen. Dit betekent dus dat 85% van de meningen zijn gegeven ten voordele van de nieuwe notatie. Het thema handen komt slechts 8 keer voor en wordt dus conform de beoordelingscriteria niet in de conclusie meegenomen. In principe geldt voor de overige elementen (i.e. competence, internal goal en value stream/importance) dat de kwalitatieve data de representatievoorkeur voor de 'nieuwe' notatie niet ondersteunen. 3 thema's kwamen te weinig in de data voor, bij 4 thema's waren te weinig contrasterende meningen.

Tabel 37: Resultaten analyse kwalitatieve feedback representatievoorkeur

Element	Thema	% thema genoemd t.o.v. totale populatie (aantal keer genoemd)	% contrasterende meningen t.v.v. nieuwe notatie (aantal keer keuze voor nieuw;oud)	Ondersteunend t.v.v. representatievoorkeur nieuwe notatie
Value proposition	Cadeau	41% (18)	83% (15;3)	ja
	Schaal	32% (14)	100% (14;0)	ja
Customer goal	Stakeholders	45% (20)	85% (17;3)	ja
	Handen	18% (8)	88% (7;1)	nee
Competence	Lamp	39% (17)	47% (8;9)	nee
	Podium	36% (16)	63% (10;6)	nee
Internal goal	Fabriek	45% (20)	60% (12;8)	nee
	Tandwielen	11% (5)	80% (4;1)	nee
Value stream/ Importance	Kleuren	48% (21)	52% (10;11)	nee
	Uitroeptekens	18% (8)	75% (2;6)	nee